Highlights

Higher-Education Expansion and Upper-Secondary Choice: RuralUrban Divergence in Vocational High School Completion

Yiqun Tong, Zhejian Wang, Ruoming Zhang

- China's 1999 tertiary expansion reshaped vocational high school attainment.
- We use 2015 nationally representative census microdata with cohort exposure.
- Cohort-based difference-in-differences implies a 3 pp rural–urban widening.
- Event studies show no pre-trends and persistent post-reform divergence.
- Effects are stronger where pre-reform higher-education density is higher.

Higher-Education Expansion and Upper-Secondary Choice: RuralUrban Divergence in Vocational High School Completion

Yiqun Tong[a], Zhejian Wang[b,*], Ruoming Zhang[b]

[a]Department of Economics, Pennsylvania State University,
[b]Department of Economics, University of Delaware,

Abstract

This paper examines whether rapid higher-education expansion reshapes earlier educational sorting in tracked systems. Using China's 1999 expansion as the setting, we analyse nationally representative microdata from the 2015 1% Population Sample Survey, which distinguishes academic senior high school from secondary vocational schooling. We exploit cohort-based exposure by comparing birth cohorts that made upper-secondary schooling decisions before the reform with cohorts young enough to adjust afterward, and we estimate difference-in-differences models that contrast rural and urban trends while controlling flexibly for geography. Across specifications, exposed rural cohorts become more likely than their urban counterparts to end schooling at vocational high school. The implied rural–urban widening in vocational high school completion as the highest attained level is about 3 percentage points in our baseline estimates. Event-study results show no comparable pre-reform trend and indicate that the divergence appears with the first exposed cohorts and remains elevated for later cohorts. Overall, the evidence suggests that mass tertiary expansion can generate sizable upstream adjustments within upper-secondary education, with consequences for the composition of terminal schooling in developing and middle-income contexts.

Keywords: Higher education expansion, Vocational education, Track choice, Ruralurban disparities, China

1. Introduction

Secondary vocational education is a central pathway in many developing countries for students who do not remain on an academic track (Bennell, 1996; Krafft, 2018). In China, vocational high schools have long played this role, but rural and urban students face starkly different information environments, household constraints, and access to academic-track capacity. These disparities imply that a common national shock to schooling incentives can generate uneven responses across space, reshaping who ends up with vocational high school as the terminal credential. China's 1999 higher-education expansion provides such a shock (Jia, Khanna, Li, & Xu, 2025; Ou & Hou, 2019; State Council of the People's Republic of China, 1999; Wan, 2006). This paper asks a focused, policy-relevant question: Did the 1999 expansion widen the rural–urban gap in vocational high school attainment?

The expansion rapidly increased university enrolment nationwide and, in doing so, raised the expected payoff to remaining on the academic route. More broadly, our question fits the canonical human-capital view that schooling choices respond to expected returns, while the incidence of return shocks can depend on frictions and constraints (Card, 1999; Goldin & Katz, 2018; Pyatt, 1966). Existing research largely evaluates this reform through college enrolment and tertiary outcomes (F. Dai, Cai, & Zhu, 2022; Huang, Tani, Wei,

*Corresponding author. Email: <zhejianw@udel.edu>

& Zhu, 2022; Jia et al., 2025; S. Li, Whalley, & Xing, 2014; Ou & Hou, 2019; Wan, 2006). We instead examine an upstream and understudied margin: whether the reform changed the likelihood that vocational high school becomes the terminal level of schooling, and whether this change differed between rural and urban cohorts. Put differently, rather than asking whether education rose overall, we ask whether the 1999 expansion reallocated terminal upper-secondary outcomes in a way that widened the rural–urban disparity in vocational attainment.

We motivate rural–urban divergence in vocational attainment with three mechanisms that all operate directly on this vocational margin. First, information and expectation updating may differ: urban families may learn more quickly that the expansion increases the value of pursuing longer academic pathways, reducing the appeal of vocational high school as a terminal choice; rural families may update more slowly or with more uncertainty. Second, household constraints may bind more tightly in rural areas. Even if expected returns to longer schooling rise, the direct costs and opportunity costs of extending education can be harder to finance, making vocational high school relatively more likely to remain the endpoint. Third, local capacity and feasibility differ: urban areas typically have better access to academic senior high schools and higher-quality schools, which can facilitate switching away from vocational options; rural areas may face limited academic capacity, making vocational schooling a more common terminal outcome even when incentives change. These mechanisms yield the same empirical prediction: after the expansion, the vocational-high-school endpoint may become relatively more prevalent in rural cohorts than in urban cohorts, widening the rural–urban gap in vocational attainment.

We test this prediction using nationally representative microdata from China's 2015 1% Population Sample Survey, which identifies vocational high school attainment separately from other schooling categories. Our empirical strategy exploits cohort-based exposure to the 1999 expansion: birth cohorts whose upper-secondary decisions were largely completed before the reform serve as a comparison group, while younger cohorts who could adjust their schooling choices after the expansion form the exposed group. We implement a rural–urban difference-in-differences design with rich geographic fixed effects, and we complement it with event-study specifications that trace how the rural–urban gap evolves across cohorts around the first exposed cohorts. This approach is designed to isolate changes in rural–urban differences in vocational high school attainment that coincide with the reform timing.

The results show a clear and economically meaningful divergence. After the higher-education expansion, vocational high school becomes relatively more common as a terminal credential for rural cohorts than for urban cohorts. In the baseline specification, the post-expansion increase in the rural–urban gap in vocational high school attainment is approximately three percentage points. Dynamic estimates indicate that this divergence emerges with the first exposed cohorts and remains elevated thereafter, while pre-exposure cohorts display no comparable differential movement. In other words, the expansion is associated with a persistent shift in the rural–urban difference in vocational outcomes, rather than a short-lived cohort-specific fluctuation.

We further examine heterogeneity patterns to characterize where the divergence is strongest. The rural–urban widening is larger in areas with stronger pre-period educational capacity and in provinces with denser pre-reform higher-education infrastructure, consistent with the idea that local feasibility and informational environments condition how quickly households respond to a nationwide reform. Importantly, the heterogeneity evidence also points to a potential role for information channels. While we find limited systematic steepening with measured provincial expansion intensity (scaled by the size of the College Entrance Examination examinees), the ruralurban divergence varies significantly with provincial differences in higher-education densityespecially the density of four-year universities. This contrast is consistent with the notion that information about the local presence and quality of universities is more salient and easier for students and families to observe than policy-induced changes in cohort-specific seat allocations, and that such salient signals may differentially shape rural and urban responses. Taken together, these findings suggest that higher-education expansion can have unintended distributional consequences at the vocational margin: even when the reform is nationwide, the vocational endpoint can move in opposite directions across rural and urban cohorts, increasing spatial inequality in educational attainment within the vocational sector.

This paper makes three main contributions. First, it documents a previously overlooked distributional consequence of tertiary massification in a tracked system: China's 1999 higher-education expansion is associated with a widening rural–urban gap in vocational high school attainment, a key terminal upper-secondary outcome (Ou & Hou, 2019; Wan, 2006; Zhang, 2026). By focusing on vocational high school as the endpoint, we show that nationwide higher-education reforms can reshape inequality not only at the university margin but also in who sorts into vocational schooling across space (Jia et al., 2025; Ou & Hou, 2019; Wan, 2006).

Second, the paper provides evidence on where this divergence is most pronounced, linking the widening vocational gap to pre-existing differences in local educational environments. The rural–urban divergence is larger in areas with stronger baseline educational capacity and in provinces with denser pre-reform higher-education infrastructure, highlighting how local constraints and informational environments mediate the incidence of a common national reform (Anderson, King, & Wang, 2003; Filmer, 2007; Ray & Ruseski, 2026).

Third, the findings speak directly to development policy debates about the role of vocational education in middle-income countries. They suggest that expanding higher educationwhile increasing opportunities at the topmay simultaneously increase spatial inequality in vocational endpoints unless complementary policies address rural constraints and upper-secondary capacity. This perspective reframes vocational high school not merely as a fallback option, but as a margin where large-scale education reforms can generate unintended distributional effects (Chen, Fu, & Pan, 2019; Liang, 2001).

The remainder of the paper is organized as follows. Section 2 describes the institutional background of vocational high schools and the 1999 expansion. Section 3 introduces the data and measures of vocational attainment and rural status. Section 4 presents the empirical design. Section 5 reports baseline estimates and event-study dynamics, and Section 6 explores heterogeneity and interpretation. Section 7 concludes.

2. Institutional Background

2.1. The 1999 Higher-Education Expansion

In January 1999, the State Council issued Notice No.4, circulating the Ministry of Education's *Action Plan for Invigorating Education toward the 21st Century* and setting a goal of raising the gross higher-education enrolment rate to about 15% by 2010 (State Council of the People's Republic of China, 1999). Policymakers presented the expansion as part of a national human-capital strategyrapidly scaling up tertiary capacity to broaden access to higher education and to support China's move toward a more knowledge-intensive economy (State Council of the People's Republic of China, 1999). Implementation relied primarily on centrally set admission quotas: first-year intake at regular higher-education institutions increased from 1.08 million in 1998 to 1.59 million in 1999 (a 47.4% increase), marking the start of a sustained enrolment surge (Fan & Ye, 2021).

Yet several years into the expansion, cohort-level access to tertiary education remained limited: in 2003, only 25% of 18-year-olds participated in the college entrance exam (CEE) and 17% accessed any college

(5% entered four-year institutions), while elite-college access was much rarer (2% for Project 211 and 1% for Project 985, and only 0.03% for *Peking University* or *Tsinghua University*) (H. Li, Loyalka, Rozelle, Wu, & Xie, 2015). Importantly, inequality arises already at the exam-participation margin: only about 12% of rural youth from nationally designated poor counties took the CEE compared with 67% of urban youth, and their cohort-level chances of accessing any college (7% vs. 48%) and Project 211 colleges (0.6% vs. 7%) were dramatically lower (H. Li et al., 2015).

2.2. Upper-Secondary Tracking in China: Academic vs Vocational

China's senior secondary education is organised as a tracked system that channels students into distinct upper-secondary pathways after completing nine years of compulsory schooling. In institutional terms, senior secondary schooling comprises a academic stream and multiple forms of secondary vocational education, including technical and specialized secondary schools, vocational senior secondary schools, and craftsmen schools. A salient feature of this system is that sorting largely occurs at the end of Grade 9 and is closely tied to exam-based selection and the allocation of seats across school types (Chen et al., 2019; Liang, 2001).

The academic track is designed and widely perceived as the primary conduit to tertiary education. In a detailed institutional overview, Liang (2001) notes that general senior secondary schools follow an academic curriculum with university entrance as the central objective, and are the most popular option among students and parents precisely because they are viewed as the main channel that may lead to higher education. In contrast, secondary vocational schooling is commonly treated as a terminal credential in practice. While 'vertical' progression routes exist on paper, they account for a small minority, and the modal transition from secondary vocational programs is direct labour-market entry rather than progression to junior college or university.[1] Moreover, entry into regular tertiary programs remains predominantly organized around the academic senior-secondary curriculum and the national college entrance pathway, implying that most junior-college entrants are drawn from general-track cohorts rather than from secondary vocational streams. Conceptually, this institutional structure makes secondary vocational schooling both a key sorting node at the upper-secondary transition and, for the vast majority of students, the effective endpoint of the vocational track.

In contrast, the secondary vocational track is institutionally framed as an employment- and skills-oriented pathway. The *Vocational Education Law* defines vocational education as a component of the national education system intended to enhance workers' skills and promote economic development, and it emphasises integration of schooling with production and close links with enterprises (Standing Committee of the National People's Congress of the People's Republic of China, 1996). International policy reviews similarly characterise upper-secondary vocational programmes as oriented toward occupational skills formation and workplace training engagement (Kuczera & Field, 2010).

Historically, vocational institutions differed not only in curriculum content but also in their perceived school-to-work linkage. For example, technical and specialized secondary schools were traditionally used to train technicians for government departments and were associated with job assignment and employment guarantees, which were later abolished amid labour-market reforms in the mid-1990s (Liang, 2001). Vocational high schools were described as offering programmes aligned with direct labour-market entryoften in expanding service-sector occupationsmaking them attractive to some students in the pre-expansion period due to the prospect of relatively immediate employment (Liang, 2001).

[1]Consistent with this characterization, a Ministry of Education sample survey reported by CCTV.com (2002) notes that 78.4% of secondary vocational school graduates were employed, and less than 10% continued further study.

Taken together, the dual-track system implies that senior-secondary school type is not merely a label but a bundle of pathways: (i) different curricular emphases (academic preparation versus occupation-specific training), (ii) different expected transitions (tertiary progression versus direct labour-market entry), and (iii) an entrenched status hierarchy and limited permeability between tracks. Consistent with this view, policy research highlights that the way tracking operates between academic and vocational streams can shape students' decisions and subsequent opportunities, and it is often discussed as a key feature of upper-secondary expansion (Chen et al., 2019).[2]

2.3. RuralUrban Inequality in Upper-Secondary Schooling

In China, the rural–urban divide in post-compulsory schooling is institutionalized through both the household registration (hukou) hierarchy and a rural–urban school system that allocates resources unevenly across locations (Hao, Hu, & Lo, 2014).

Consistent with this institutionalization, evidence on post-expansion higher-education access suggests that observed gaps are driven primarily by rural–urban differences rather than by differences between poor and non-poor counties within rural or urban areas (H. Li et al., 2015). National and World Bank assessments document persistent upper-secondary access gaps: only 77 percent of youth have completed high school by age 20, and the completion rate for rural youth is about 47 percent (Chen et al., 2019). These statistics pertain to the post-2015 period; the corresponding rural–urban gap was markedly larger around 1999, reflecting much more unequal access to upper-secondary schooling and, consequently, to the high-school-to-college transition. These disparities also reflect systematic differences in guidance and information; for example, limited career counseling leaves middle-school students unprepared to make a well-informed Grade-9 track choice (Chen et al., 2019). Microdata evidence indicates large hukou-based gaps in progression from junior high to both academic and vocational senior high school (Yeung, 2013). Rural households face systematically higher risks and constraints in pursuing the academic track at the secondary level.

3. Conceptual Model

3.1. Model Structure

A unit mass of junior-secondary graduates is indexed by i. Each student belongs to a group $g \in \{U, R\}$ (urban vs. rural). Let λ_g denote the mass of group g, with $\lambda_U + \lambda_R = 1$. Each student has a type: $\theta_i = (a_i, W_i)$, where $a_i \in \mathbb{R}_+$ represents the ability, and $W_i \in \mathbb{R}_+$ represents the family resources. We interpret W_i as a reduced-form measure of family wealth, borrowing capacity, and effective parental support.

Within each group, (a, W) is distributed according to a distribution $F_g(a, W)$. We further assume that rural family resources are lower in the sense of first-order stochastic dominance: letting $F_g^W(\cdot)$ denote the marginal distribution of W in group g, we have $F_R^W(w) \geq F_U^W(w)$ for all w, implying rural households are more likely to be resource-constrained.

After junior secondary school, each student chooses a route $s_i \in \{G, V, O\}$, where G denotes the academic high school track, which keeps the possibility to go for higher education; V denotes the vocational high school track, which we treat as a terminal with no access to further schooling; and O denotes exit from schooling after junior secondary school.

[2] For evidence that vocational schooling in China is associated with distinct labour-market returns (and is therefore substantively meaningful as an educational category rather than a purely administrative label), see L. Dai and Martins (2024).

Present value lifetime incomes for different education levels are $y_O(a), y_H(a), y_V(a), y_C(a)$, corresponding to exit, academic high school without college, vocational high school, and college.

Let (N_G, N_V) denote the aggregate masses choosing G and V:

$$N_G \in [0,1], \qquad N_V \in [0,1], \qquad N_O = 1 - N_G - N_V.$$

Each route requires a cost that depends on aggregate participation:

$$I_g^G(N_G) \equiv \kappa_g^G + c^G(N_G), \qquad I_g^V(N_V) \equiv \kappa_g^V + c^V(N_V),$$

where $\kappa_R^G > \kappa_U^G \geq 0$ and $\kappa_R^V > \kappa_U^V \geq 0$ capture higher costs for rural households in both tracks—e.g., longer distances, boarding needs, and higher out-of-pocket and time costs to keep a student in high school.

We assume $c^G(\cdot)$ and $c^V(\cdot)$ are continuous and increasing:

$$\frac{dc^G(N_G)}{dN_G} > 0, \qquad \frac{dc^V(N_V)}{dN_V} > 0.$$

Intuitively, $c^G(N_G)$ and $c^V(N_V)$ capture reduced-form congestion and competition costs: larger cohorts strain school resources (a soft capacity constraint) and intensify competitive admissions, requiring greater effort and expenditures to enter and complete each track.

If family resources cannot cover the high school cost, the route is infeasible. Specifically, route $j \in \{G, V\}$ is feasible for student i if and only if $W_i \geq I_g^j(N_j)$.

Let $K > 0$ be the number of higher-education admissions (i.e., college seats) available for the cohort. If a student chooses G, the probability of college admission is

$$p_i \equiv p(a_i; K, N_G).$$

We assume $p(a; K, N)$ is continuous and satisfies:

$$\frac{\partial p}{\partial a} > 0, \qquad \frac{\partial p}{\partial K} > 0, \qquad \frac{\partial p}{\partial N} < 0.$$

College cost is $C \geq 0$ (tuition, living expenses, foregone earnings in present value). This cost is incurred only if the student enrolls in college. Conditional on admission, college enrolment requires financing the college cost. Let $\pi(\cdot)$ denote the probability of financing college enrolment as a function of remaining resources after paying the high-school investment:

$$\pi_i \equiv \pi\big(W_i - I_g^G(N_G)\big) \in [0,1].$$

We assume that $\pi(\cdot)$ is continuous and increasing with $\lim_{x \to \infty} \pi(x) = 1$.

Define the effective probability of completing the general-to-college route:

$$q_i \equiv p(a_i; K, N_G) \cdot \pi\big(W_i - I_g^G(N_G)\big)\,.$$

All students share the same increasing, strictly concave utility function $u : \mathbb{R}_+ \to \mathbb{R}$, with $u'(c) > 0, u''(c) < 0$.

We have the utilities for the three different tracks below:

$$U_i(O) = u(W_i + y_O(a_i))\,.$$

$$U_i(V \mid N_V) = u\big(W_i - I_g^V(N_V) + y_V(a_i)\big)\,.$$

$$U_i(G \mid N_G) = q_i \cdot u\big(W_i - I_g^G(N_G) + y_C(a_i) - C\big) + (1 - q_i) \cdot u\big(W_i - I_g^G(N_G) + y_H(a_i)\big)\,.$$

Given (K, N_G, N_V), each student chooses a feasible route maximizing utility:

$$s_i \in \arg\max \Big\{U_i(O),\ U_i(V \mid N_V) \text{ if feasible},\ U_i(G \mid N_G) \text{ if feasible}\Big\}.$$

Students are nonatomic and take (N_G, N_V) as given.

3.2. Equilibrium

Let $BR(a, W; g, K, N_G, N_V)$ denote the best-response correspondence.

Definition 1 (Track-choice equilibrium)**.** *An equilibrium is a pair (N_G^*, N_V^*) with $N_G^*, N_V^* \geq 0$ and $N_G^* + N_V^* \leq 1$ such that, when all students best respond to (K, N_G^*, N_V^*), the induced aggregate masses choosing G and V equal (N_G^*, N_V^*).*

Define the aggregate mappings:

$$\Phi_G(N_G, N_V; K) = \sum_{g \in \{U,R\}} \lambda_g \int \mathbf{1}\{G \in BR(a, W; g, K, N_G, N_V)\}\, dF_g(a, W),$$

$$\Phi_V(N_G, N_V; K) = \sum_{g \in \{U,R\}} \lambda_g \int \mathbf{1}\{V \in BR(a, W; g, K, N_G, N_V)\}\, dF_g(a, W).$$

Equilibrium (N_G^*, N_V^*) satisfies

$$N_G^* = \Phi_G(N_G^*, N_V^*; K), \qquad N_V^* = \Phi_V(N_G^*, N_V^*; K).$$

Under some standard assumptions, a unique equilibrium exists.

3.3. Empirical Implications

3.3.1. Higher-education expansion increases the relative appeal of the academic track

Holding congestion fixed (i.e., holding N_G fixed), an increase in K raises the college-admission probability $p(a; K, N_G)$, thus increasing the expected utility of choosing G for students among those who can access high school in the first place. Therefore, in both urban and rural areas, some students who previously chose vocational high school switch to the general track.

3.3.2. Vocational enrolment declines in urban areas

In urban areas, the baseline mass choosing O is small. As a result, the main adjustment after the expansion occurs within high school, predominantly through $V \to G$. Even if some entry into V occurs, the thin urban O margin implies that such entry cannot offset the $V \to G$ move-out. Hence, the net change in urban vocational enrolment is negative.

3.3.3. Vocational enrolment rises in rural areas

Vocational costs depend on aggregate vocational participation. Since urban vocational enrolment is a large component of total vocational enrolment, the urban shift $V \to G$ reduces aggregate N_V, which lowers vocational cost $I_R^V(N_V)$ faced by rural households. This makes vocational high school newly feasible or attractive for some rural students who previously chose O, generating rural entry $O \to V$.

Because rural areas have a relatively larger pool of students at O and a smaller baseline vocational share, the entry effect can dominate the rural move-out from $V \to G$, so rural vocational enrolment can increase in equilibrium.

4. Data

4.1. 2015 1% Population Sample Survey of China

We use microdata from the 2015 1% Population Sample Survey of China as our primary data source, a nationally representative microdataset administered by the National Bureau of Statistics. The survey is drawn from the full decennial census and covers roughly one percent of the national population, comprising over one million individuals. It records detailed demographic, educational, occupational, and geographic information, making it one of the most authoritative sources for studying long-run socioeconomic outcomes at the individual level. Crucially, the dataset includes information on parental education, employment industry, and land-contract arrangements, which we use to construct key variables on treatment exposure, rural versus urban status, and heterogeneous effects.

The 2015 wave provides a unique opportunity to assess the consequences of education reforms from the late 1990s. It observes completed educational attainment and early labour-market outcomes for individuals born in the 1980s, a cohort directly affected by the 1999 expansion of higher education. Crucially, it is the only census-scale dataset that reports full schooling outcomes while distinguishing academic from vocational programmes at the upper-secondary level.

We restrict our sample to individuals born between 1970 and 1988, a 19-cohort window that spans the pre- and post-reform births relevant for our design. Consistent with the standard age-at-entry rule, students sitting for the 1999 college entrance examination were typically 18 or 19 years old, corresponding roughly to the 1981 or 1980 birth cohorts. Including earlier birth cohorts would add many observations but little additional identifying variation, because their schooling and initial labour-market decisions were completed well before the 1999 expansion and under different institutional conditions. Cohorts at the upper end of our window (1970–1979) had already completed high school by the time the June 1999 university expansion was announced and therefore serve as an unaffected baseline, while cohorts at the lower end (1980–1988) were still in secondary school or approaching the high-school-to-college transition and thus could adjust their upper-secondary track choices and subsequent postsecondary plans in response to the reform. We therefore take 1979 as the base year: it is safely pre-reform and, by construction, belongs to the youngest cohorts whose high-school completion and exam timing were largely settled prior to the 1999 expansion. Extending the window to cohorts born after 1988 would both compress the time available to observe stable labour-market outcomes by 2015 and increase potential contamination from subsequent education reforms, notably the 2006–2007 abolition of rural compulsory-education fees in rural areas.[3] After dropping observations with missing covariates, the working sample comprises around 400,000 individuals.

[3]The 2006–2007 reform eliminated tuition and miscellaneous fees for compulsory education in rural areas, which could differentially affect later cohorts' schooling decisions and confound the interpretation of post-1999 dynamics.

4.2. Key Variables

Our empirical strategy measures educational attainment using credential-based outcomes that place individuals at distinct points along the schooling hierarchy, with particular emphasis on sorting at the upper-secondary stage and the subsequent transition into tertiary education. The 2015 1% Population Sample Survey is well suited for this purpose because its education classification distinguishes the academic upper-secondary track from the secondary vocational track. The vocational category encompasses vocational high schools as well as specialized secondary and technical schools. Exploiting this detail, we construct mutually exclusive indicators for an individuals highest qualification as of 2015, differentiating between academic high school completion, vocational high school completion[4], junior college, and university. To describe attainment at successive continuation margins, we also construct 'at least' indicators for each of these credentials, coded as one if the individual has completed that level or any higher level.

Because the census does not record hukou status directly, we proxy rural registration using land-contract rights: we classify a household as rural if it holds a long-term contract to collectively owned village farmland, and as urban otherwise. Under China's land system, eligibility for long-term contracts to collectively owned village farmland is institutionally tied to rural registration, so land-contract rights serve as a tight proxy for rural registration. Table 1 reports descriptive statistics for the main analysis sample, stratified by rural status and gender and separately for pre- and post-reform birth cohorts.

[4]In practice, vocational high school completion is a close proxy for vocational-track enrolment at the upper-secondary transition because the vast majority of secondary vocational students do not progress to tertiary education. Contemporary policy reporting indicates that only about 10% of secondary vocational graduates continued further study, with most transitioning directly into employment (CCTV.com, 2002).

Table 1: Descriptive Statistics

	Rural		Urban	
	Male	Female	Male	Female
Panel A: Pre cohorts (1970–1979)				
Years of schooling	8.799	8.211	11.835	11.121
	(2.180)	(2.334)	(3.171)	(3.293)
Vocational high school	0.020	0.013	0.096	0.089
	(0.142)	(0.114)	(0.294)	(0.285)
Han ethnicity	0.893	0.897	0.934	0.928
	(0.309)	(0.304)	(0.248)	(0.259)
General HS	0.119	0.072	0.525	0.436
	(0.323)	(0.258)	(0.499)	(0.496)
Higher education	0.021	0.011	0.323	0.253
	(0.142)	(0.103)	(0.468)	(0.435)
N	71,110	61,683	38,384	42,014
Panel B: Post cohorts (1980–1988)				
Years of schooling	9.912	9.621	12.527	12.150
	(2.533)	(2.586)	(3.264)	(3.402)
Vocational high school	0.052	0.048	0.092	0.086
	(0.223)	(0.214)	(0.290)	(0.280)
Han ethnicity	0.894	0.893	0.908	0.913
	(0.308)	(0.309)	(0.288)	(0.282)
General HS or above (share)	0.248	0.205	0.594	0.543
	(0.432)	(0.404)	(0.491)	(0.498)
Higher education	0.096	0.084	0.435	0.399
	(0.294)	(0.278)	(0.496)	(0.490)
N	61,298	52,169	35,964	42,189

Notes: The table reports means with standard deviations in parentheses. Cohorts are defined by birth year: pre cohorts are 1970–1979 and post cohorts are 1980–1988. Rural/urban status is based on hukou type. 'Vocational high school' is an indicator for individuals with vocational high school as the highest educational attainment; its mean is the vocational high school share. 'General HS or above' equals one for academic high school or higher education (college or above), excluding vocational high school. 'Higher education' equals one for college (associate) or above. Han ethnicity is an indicator for Han. Years of schooling follows a coarse mapping from the education-category variable.

5. Empirical Strategy

5.1. Identification Strategy and Cohort Exposure

We first estimate the average treatment effect of the expansion on rural students relative to their urban counterparts using the following linear probability model:

$$Y_{icb} = \alpha + \beta(Rural_i \times Post_b) + \Gamma X_{ic} + \delta_c + \gamma_b + \theta_{cr} + \epsilon_{icb} \quad (1)$$

where Y_{icb} is an indicator equal to 1 if individual i in county c from birth cohort b *completes vocational high school*.[5] $Post_b$ equals 1 for cohorts exposed to the 1999 higher-education expansion—defined as cohorts

[5]In the empirical implementation, Y_{icb} corresponds to the binary outcome `vocational_high`.

born after 1980—and 0 otherwise, and $Rural_i$ is a binary indicator for rural status at the time of the survey. To clarify how the 1999 expansion maps into cohorts' schooling stages, Appendix Table A.1 provides an illustrative cohort-to-grade mapping around June 1999. The coefficient of interest, β, captures the differential post-policy change in the probability of vocational high school completion for rural individuals relative to urban individuals. X_{ic} is a vector of individual-level controls, including gender and ethnicity. The specification includes county fixed effects δ_c, birth-cohort fixed effects γ_b, and county-by-rural fixed effects θ_{cr}. Standard errors are clustered at the county level to account for within-county correlation in the error term ϵ_{icb}.

5.2. Event-Study Design

The validity of our DiD estimator relies on the parallel trends assumption—that in the absence of the 1999 expansion, the gap between rural and urban vocational high school completion would have remained constant. To test this and examine the dynamic evolution of the gap, we estimate an event study specification:

$$Y_{icb} = \alpha + \sum_{k=1970, k\neq 1979}^{1988} \beta_k (\mathbb{1}\{BirthYear = k\} \times Rural_i) + \Gamma X_{ic} + \delta_c + \gamma_b + \theta_{cr} + \epsilon_{icb} \tag{2}$$

In this model, we interact the rural dummy with a full set of birth cohort indicators. We omit the year **1979** as the reference category ($k = 1979$). Consequently, each β_k represents the urban-rural gap in vocational high school completion for cohort k relative to the gap that existed in the 1979 baseline cohort. Evidence of parallel trends is supported if β_k is statistically insignificant for all pre-treatment cohorts.

5.3. Geographic Variation in Exposure: Baseline Local Educational Capacity

The effects of the 1999 higher-education expansion may plausibly vary across places with different pre-existing educational capacity. To study this geographic heterogeneity in a predetermined way, we focus on birth cohorts from 1975 to 1988 to more precisely estimate cross-county differences in the rural–urban response. We construct a county-level baseline measure of local educational capacity using only pre-expansion cohorts. Specifically, for each county, we compute the share of individuals in the 1974 birth cohort who attained any type of high school education. We then classify counties into high- versus low-capacity groups based on whether this baseline share is above or below the cross-county median. Because this grouping is defined exclusively from pre-policy cohorts, it is fixed prior to treatment and is not mechanically affected by post-expansion outcomes.

We incorporate this baseline grouping into an extended event-study (triple-difference) specification that allows the rural–urban gap to evolve differentially across high- and low-capacity counties:

$$\begin{aligned} Y_{icb} = \alpha + \sum_{k\neq 1979} \beta_k \Big(\mathbb{1}\{b = k\} \times Rural_i\Big) + \sum_{k\neq 1979} \phi_k \Big(\mathbb{1}\{b = k\} \times Rural_i \times HighCap_c\Big) \\ + \Gamma' X_i + \delta_c + \gamma_b + \theta_{cr} + \psi_{b,HighCap} + \epsilon_{icb}, \end{aligned} \tag{3}$$

where Y_{icb} is an indicator equal to 1 if individual i in county c from birth cohort b *completes vocational high school*, $Rural_i$ indicates rural status, and $HighCap_c$ denotes counties with above-median baseline educational capacity. X_i includes individual-level controls. The specification absorbs county fixed effects (δ_c), birth-cohort fixed effects (γ_b), and county-by-rural fixed effects (θ_{cr}), and additionally includes cohort-by-capacity-group fixed effects ($\psi_{b,HighCap}$) to flexibly account for cohort-specific shocks common to counties in the same baseline-capacity group. Standard errors are clustered at the county level.

The coefficients ϕ_k summarize the heterogeneity of the cohort-specific rural–urban gap in vocational high school completion across high- versus low-capacity counties, relative to the 1979 reference cohort. A positive (negative) ϕ_k for post-expansion cohorts implies that the rural–urban gap widens (narrows) more in counties with higher pre-expansion educational capacity.

6. Results

6.1. Higher-Education Expansion and Vocational High-School Completion

This subsection documents how cohort exposure to the 1999 higher-education expansion is associated with changes in upper-secondary terminal schooling, measured as an indicator for completing vocational high school as the highest attained level. The estimating variation comes from the interaction between cohort exposure (*Post*, defined as birth cohorts after 1980) and rural residence (*Rural*), so that the coefficient on $Rural \times Post$ captures the differential post-expansion change in vocational high school completion for rural individuals relative to urban individuals, conditional on the included controls and fixed effects.

In the main difference-in-differences specification, the estimated $Rural \times Post$ coefficient is 0.030 (Table 2). This corresponds to a 3.0 percentage point increase in the probability that rural individuals in exposed cohorts end schooling at vocational high school relative to the corresponding rural–urban difference among earlier cohorts. The estimate is precise under county-clustered standard errors. To make clear which group-level movements underpin this differential estimate, Figure 1 plots raw cohort means of terminal vocational high-school attainment separately for rural and urban individuals. The descriptive series suggests that the post-1980 widening in the rural–urban gap reflects both a continued increase among rural cohorts and a decline (or stagnation) among urban cohorts.

The results are similar by gender. For males, the estimated differential increase is 0.028, and for females it is 0.033, both statistically significant at the 1% level (Table 2). The magnitudes are comparable across the two subsamples.

Figure 2 plots cohort-specific estimates of the rural–urban gap in vocational high school completion, normalised to the 1979 birth cohort, with 95% confidence intervals. The event-study design serves two purposes. First, it provides a direct diagnostic for differential pre-trends: the coefficients for pre-1980 cohorts are generally small and do not display a systematic upward pattern relative to the reference cohort, offering reassurance that the post-1980 change is not simply the continuation of an earlier cohort process. Second, it characterises the timing of the post-period adjustment: the rural–urban gap increases after the first exposed cohort and remains elevated in subsequent cohorts, indicating that the average positive $Rural \times Post$ estimate in the main specification reflects a persistent (and potentially evolving) post-exposure widening rather than a one-cohort shift.

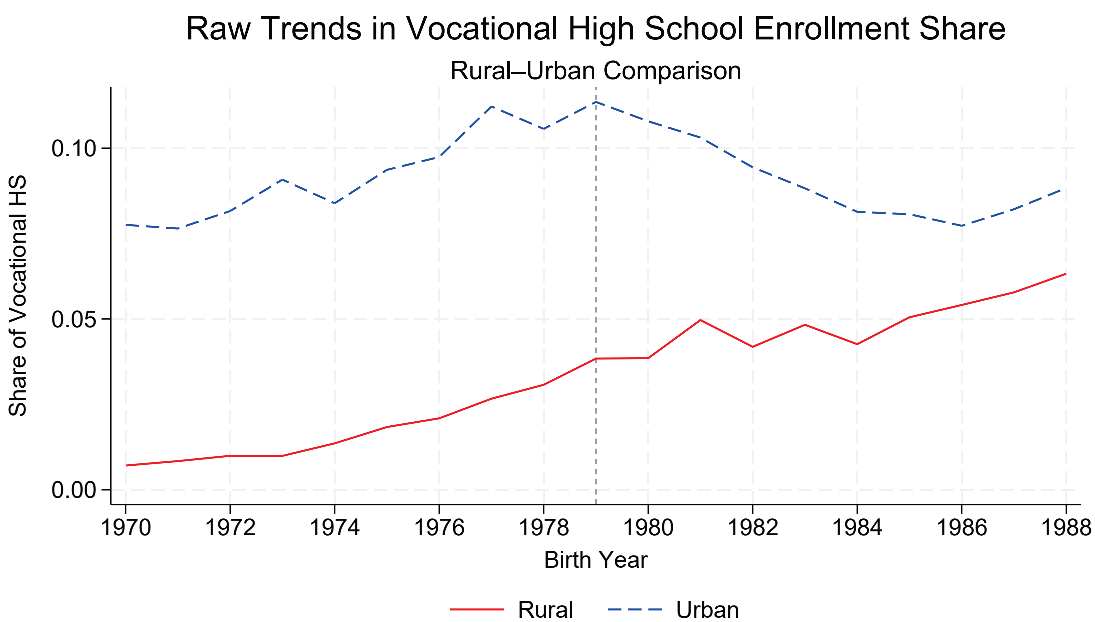

Figure 1: Raw cohort trends in terminal vocational high-school attainment by rural status

Notes: The figure plots cohort means of an indicator equal to one if the respondent's highest completed education is secondary vocational high school. Rural status follows the baseline definition in the main analysis. The dashed vertical line marks the 1979 birth cohort (the reference cohort in the event-study analysis); cohorts born after 1980 are classified as exposed ($Post = 1$). These are unadjusted descriptive trends; the regression estimates reported in Table 2 and Figure 2 condition on controls and fixed effects as described in the text.

Table 2: Higher-Education Expansion and Vocational High School Completion

	End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural × Post	0.030***	0.028***	0.033***
	(0.002)	(0.003)	(0.002)
Controls	Yes	Yes	Yes
County and Cohort FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Observations	404,782	206,671	197,979
R^2	0.057	0.071	0.069
Clustering Level	County	County	County

Notes: The dependent variable is an indicator equal to one if an individuals highest completed level of schooling is vocational high school. *Post* indicates cohorts exposed to the 1999 higher-education expansion (birth cohorts after 1980), and *Rural* is a rural-residency indicator measured at the time of the 2015 census. Robust standard errors clustered at the county level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

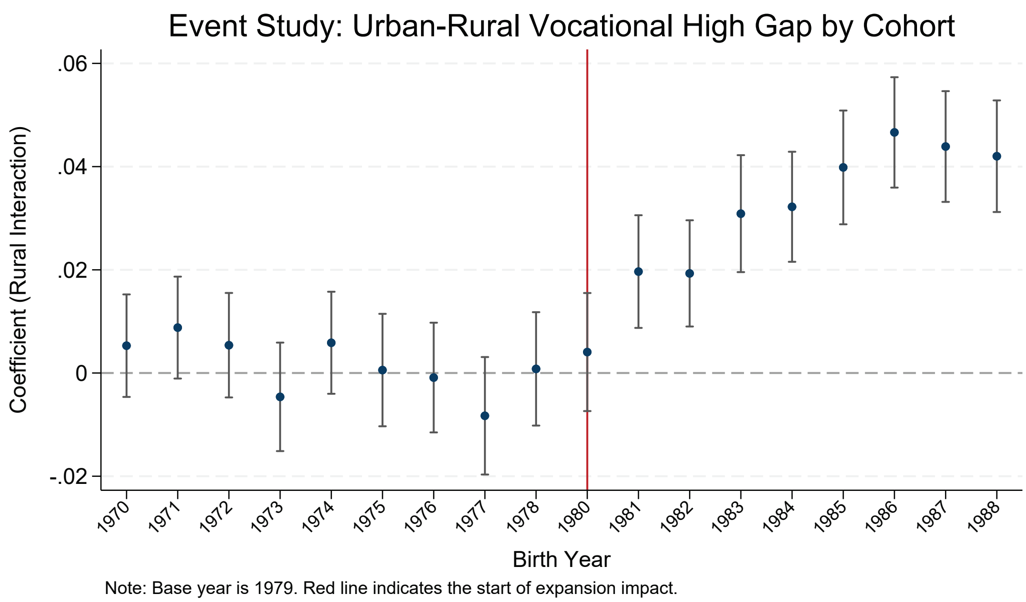

Figure 2: Event Study: Rural–Urban Gap in Vocational High School Completion

Notes: The figure plots cohort-specific estimates of the rural–urban gap in the probability that an individuals highest completed level of schooling is vocational high school, compare to the 1979 birth cohort. Points show estimated coefficients and vertical bars denote 95% confidence intervals. The vertical line marks the first cohort classified as exposed to the 1999 higher-education expansion. Standard errors are clustered at the county level.

6.2. Geographic Exposure and the RuralUrban Divergence

This subsection examines whether the rural–urban effect of the higher-education expansion varies systematically with pre-period county educational capacity. We proxy baseline capacity using a predetermined county classification based on the share attaining any high school education among the 1974 birth cohort (above- vs. below-median), and estimate a triple-difference design using birth cohorts from 1975 to 1988. For descriptive context, Appendix Figure B.1 plots raw cohort means of terminal vocational high-school attainment by rural status separately for high- and low-capacity counties. The figure illustrates that the post-expansion rural–urban divergence is more pronounced in high-capacity counties, consistent with the triple-difference estimates reported below.

The estimates indicate sizable geographic heterogeneity in the rural–urban differential response. In low-capacity counties, post-expansion cohorts exhibit a higher probability of terminating schooling at vocational high school in rural relative to urban areas by 1.9 percentage points in the pooled sample. The corresponding rural–urban differential is 1.9 percentage points for males and 1.7 percentage points for females. In high-capacity counties, this rural–urban differential is further amplified by an additional 2.5 percentage points in the pooled sample, implying a combined post-expansion rural–urban differential of 4.4 percentage points. The analogous totals are 3.3 percentage points for males and 5.4 percentage points for females, with the triple interaction precisely estimated in all three columns (Table 3). By contrast, the post-expansion change associated with high-capacity counties outside rural areas is small in magnitude, with a modest negative estimate for females.

Cohort-specific estimates of the triple interaction corroborate this pattern. The event-study series in Figure 3, normalised to the 1979 cohort, traces the difference in the rural–urban gap between high- and low-capacity counties across birth cohorts and indicates that the post-expansion increase in the rural–urban differential is concentrated among cohorts exposed to the expansion, with confidence intervals reflecting county-clustered uncertainty.

Table 3: Geographic Heterogeneity: Triple-Difference Estimates for Vocational High School Completion

	Dependent variable: End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural × Post	0.019***	0.019***	0.017***
	(0.003)	(0.005)	(0.004)
High-capacity county × Post	-0.013***	-0.004	-0.022***
	(0.004)	(0.006)	(0.005)
Rural × Post × High-capacity county	0.025***	0.014**	0.037***
	(0.005)	(0.007)	(0.006)
Controls	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	284,444	144,722	139,568
R^2	0.060	0.078	0.076
County clusters	2,716	2,708	2,709

Notes: The dependent variable equals one if an individuals highest completed level of schooling is vocational high school. 'High-capacity county' is a predetermined county classification based on the pre-period share of individuals attaining any type of high school education (above-median vs. below-median), constructed from pre-expansion cohorts. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

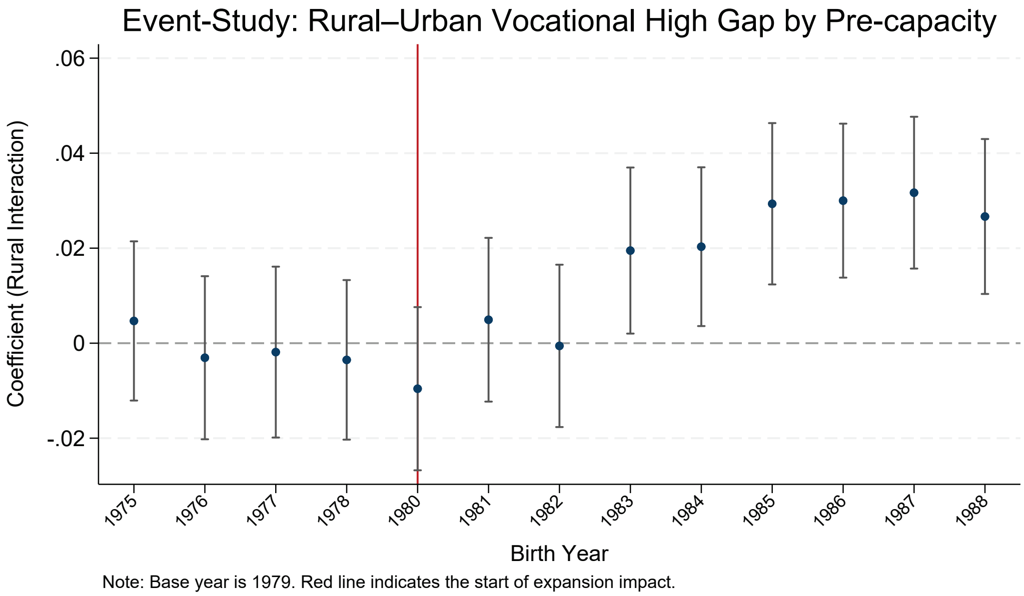

Figure 3: Event Study: Heterogeneity in the Rural–Urban Gap by Baseline County Capacity

Notes: The figure plots cohort-specific estimates of the triple interaction between the rural indicator and an indicator for above-median baseline county educational capacity, interacted with birth-cohort dummies. Coefficients are normalised relative to the 1979 cohort and can be interpreted as the difference in the rural–urban gap between high-capacity and low-capacity counties for each birth cohort (i.e., a cohort-specific triple-difference). Vertical bars indicate 95% confidence intervals. Standard errors are clustered at the county level.

7. Robustness Checks

In this section, we examine the robustness of the main Rural Œ Post estimates to a variety of alternative specifications and sample definitions. We then show that ruralurban gaps in vocational high school completion exhibit flat pre-trends prior to the reform, alleviating concerns about differential convergence that predates the expansion. Next, we verify that our results are stable when we redefine rural and urban status based on a more restrictive sample of non-migrant individuals and when we drop boundary cohorts that may only be partially exposed to the reform. Finally, we demonstrate that the estimated effects are robust to clustering standard errors at different levels, indicating that our conclusions are not sensitive to the precise outcome definition or inference procedure.

7.1. Robustness: Hukou Reclassification and Migration

A potential concern is that selective migration and hukou conversion could mechanically affect the measured rural–urban gap across cohorts. For example, if post-reform cohorts with stronger educational aspirations are more likely to move or to convert hukou status, changes in the observed rural–urban gap in the probability of terminating schooling at vocational high school could partly reflect compositional shifts rather than behavioural responses. To address this concern, we re-estimate the event-study specification restricting the sample to individuals whose hukou status did not change (Figure C.2). This restriction yields 226,133 observations across 2,831 counties. The event-study pattern remains qualitatively similar, indicating that our baseline results are not driven by hukou transfer or selective mobility.

7.2. Robustness: Excluding Urban Districts

A related concern is that the baseline rural–urban comparison may be influenced by the presence of prefecture-level city urban districts, which differ systematically from both rural counties and standard county-level urban areas in terms of labour-market conditions, educational infrastructure, and migration intensity. In particular, these districts may experience distinct schooling options or administrative arrangements that are not representative of the broader rural–urban margin of interest. To address this concern, we re-estimate the baseline DiD specification after excluding prefecture-level city urban districts and retaining only non-district counties.

As shown in Table C.2, the estimated post-expansion rural–urban differential in vocational high school completion remains very similar in magnitude and statistical significance to the baseline estimates. The results are stable in the pooled sample as well as when estimated separately by gender. This robustness check suggests that our main findings are not driven by a small set of highly urbanized administrative units and instead reflect a broader rural–urban response to higher-education expansion.

7.3. Robustness: Placebo Cohorts (19701978))

To assess whether our event-study pattern could be driven by spurious cohort trends unrelated to the 1999 expansion, we conduct a placebo test using only clearly pre-reform birth cohorts. Specifically, we restrict the sample to individuals born between 1970 and 1979 and re-estimate the same event-study specification. These cohorts completed key schooling decisions well before the June 1999 expansion was announced, so any systematic post-style dynamics within this window would cast doubt on our identifying assumptions. Reassuringly, the placebo coefficients are small and show no coherent trend across cohorts, suggesting that the baseline results are unlikely to be driven by pre-existing differential cohort trends between rural and urban areas (Figure C.3).

7.4. Robustness: Alternative Cohort Windows

A concern in cohort-based designs is that results may be sensitive to the choice of cohorts around the reform cutoff, especially if exposure is partial for cohorts close to the announcement or implementation dates. To assess this, we re-estimate the baseline specification after excluding the three cohorts closest to the 1999 expansion boundary (1979–1981). This restriction removes potentially ambiguous cohorts while keeping the fixed-effects structure unchanged. Table C.3 shows that the estimated post$\times$rural effect is very similar in magnitude and remains precisely estimated, both in the full sample and within gender subsamples.

7.5. Robustness: Alternative Clustering Choices

Our baseline inference clusters standard errors at the county level, which allows for arbitrary within-county correlation across cohorts. As a robustness check, we re-estimate the same specification but compute standard errors clustered at higher geographic levels—the province and the prefecture-level city. Clustering at higher levels is more conservative if shocks are spatially correlated beyond counties (e.g., province-wide policy implementation or city-level labour-market conditions). Table C.4 shows that the estimated effect is very similar across clustering choices and remains statistically significant in the full sample as well as within gender subsamples. This suggests that our main conclusions are not sensitive to the choice of clustering level.

8. Potential Confounders and Concurrent Shocks

Because the 1999 higher-education expansion coincided with several late-1990s macro shocks and policy initiatives that were unevenly distributed across space, a natural concern is that the estimated post-cohort rural–urban divergence in vocational high school attendance might reflect these contemporaneous changes rather than the expansion itself. To assess this, we implement a set of differential-exposure tests that allow the post-cohort rural–urban gap to vary with pre-determined measures of local exposure to major shocks and campaigns. Across these exercises, the estimated coefficient on *Rural* × *Post* remains positive and of similar magnitude (roughly 0.026–0.036 across pooled specifications), and the additional interaction terms are comparatively small, indicating that no single contemporaneous factor can account for the timing and magnitude of the observed break.

8.1. Asian Financial Crisis Exposure (1998)

The 1998 Asian Financial Crisis could plausibly affect upper-secondary track choice by changing local labour-market conditions and household resources, particularly in provinces more exposed to external trade and foreign direct investment (FDI). Table D.5 addresses this concern by splitting provinces into high- versus low-exposure regions and augmenting the baseline specification with *High Region* × *Post* and *Rural* × *Post* × *High Region*. The estimated post-cohort rural–urban divergence remains stable: the pooled *Rural* × *Post* coefficient is 0.036, and the implied rural–urban gap in high-exposure provinces is only slightly smaller (0.036–0.007 = 0.029). The same pattern holds by gender: for males, the implied high-exposure gap is 0.025+0.003 = 0.028; for females, it is 0.049–0.019 = 0.030. Overall, differential crisis exposure does not overturn the main post-cohort rural–urban divergence.

8.2. Flood Exposure (1998)

Large-scale flooding in 1998 could mechanically affect schooling decisions through local economic disruption and damage to public infrastructure, with potentially different implications for rural versus urban households. Table D.6 tests this channel using a county-level indicator for flood-affected areas and includes *Flood* × *Post* and *Rural* × *Post* × *Flood*. The estimated *Rural* × *Post* coefficient remains essentially unchanged (0.030 pooled), while the triple interaction is small relative to the baseline estimate and not systematically different from zero. This suggests that exposure to the 1998 flood shock does not explain the post-cohort rural–urban divergence in vocational attendance.

8.3. Compulsory Education Campaigns

A further concern is that contemporaneous education policy could shift the pool of students eligible for upper-secondary tracks independently of higher-education expansion. Table D.7 examines exposure to the Nine-Year Compulsory Education catch-up campaign using a county-level indicator and interactions with the post-cohort and rural status. The estimated rural–urban post-cohort divergence remains close to the baseline magnitude (0.031 pooled). The additional interaction *Rural* × *Post* × *Campaign* is modest and does not materially alter the estimated gap, indicating that differential rollout of the catch-up campaign is unlikely to drive the main pattern.

8.4. State-Owned Enterprise Restructuring

Late-1990s SOE (state-owned enterprise) restructuring could confound the estimates if provinces with larger state-sector employment experienced sharper labour-market dislocation that differentially affected rural and urban schooling decisions. Table D.8 addresses this concern by interacting the post-cohort indicator with a standardised measure of provincial SOE employment share (measured in 2000) and allowing the rural–urban post-cohort gap to vary with this pre-exposure intensity. The estimated $Rural \times Post$ coefficient remains stable (0.031 pooled). While $SOE\ share \times Post$ is negative in level terms, the incremental rural differential associated with SOE intensity is small (0.004 pooled; 0.006 for males; approximately zero for females) relative to the baseline divergence. Thus, variation in SOE restructuring exposure does not account for the post-cohort rural–urban break.

8.5. China Trade Shock

Finally, China's WTO accession and the subsequent trade shock could plausibly affect upper-secondary track choice by changing the relative returns to different skills and the composition of local labour demand, with effects that vary with trade exposure. Table D.9 incorporates a continuous measure of exposure based on per-capita trade volume and allows both the post-cohort level shift and the post-cohort rural–urban gap to vary with this measure. The estimated $Rural \times Post$ coefficient remains close to the baseline (0.026 pooled), and the triple interaction with WTO exposure is small in magnitude (0.002 pooled; 0.001 for males; 0.003 for females). These results indicate that differential exposure to trade integration does not materially change the estimated post-cohort rural–urban divergence in vocational high school attendance.

9. Heterogeneity Analysis

This section assesses whether the rural–urban divergence in upper-secondary track choice associated with the 1999 higher-education expansion varies systematically across pre-determined, policy-relevant contexts. We focus on three dimensions measured prior to the reform: (i) provincial expansion intensity, (ii) baseline higher-education availability, and (iii) regional agricultural dependence. Empirically, we augment the baseline specification by interacting the core $Rural \times Post$ term with each pre-reform provincial characteristic while maintaining the same fixed effects structure and county-level clustering as in the baseline estimates. Throughout, the emphasis is on whether the ordering and magnitude of effects are consistent with differential exposure or local capacity.

9.1. Heterogeneity by Expansion Intensity

Provincial expansion intensity is a natural ex ante source of heterogeneity because it captures how strongly a province was exposed to the higher-education expansion in the post-reform period. We construct a standardised provincial measure of expansion intensity and report its cross-province variation in Table E.10. Table 4 implements this by interacting $Post$ with this measure and adding the corresponding triple interaction $Rural \times Post \times Expansion$.

Across pooled and gender-specific specifications, the estimated $Rural \times Post$ effect remains positive and tightly estimated (0.031 in the pooled sample; 0.029 for males; 0.033 for females), indicating that the baseline rural–urban divergence in vocational high-school enrolment is not sensitive to allowing exposure to vary with measured expansion intensity. The interaction $Expansion \times Post$ is positive and statistically significant in the pooled and male samples, suggesting somewhat larger post-reform changes in provinces with higher expansion intensity, but the key differential term, $Rural \times Post \times Expansion$, is small in magnitude

(0.002 in the pooled sample) and imprecisely estimated. Overall, these results indicate that the rural–urban divergence observed in the baseline estimates is broadly present across provinces, with limited evidence that it systematically steepens with measured expansion intensity.

A key distinction is that our provincial expansion-intensity measure captures the magnitude of quota growth (intensity), which may not have been salient to rural households at the time of track choice, whereas baseline local university density captures the visibility and perceived accessibility of higher education (visibility), and thus is more likely to generate systematic heterogeneity in rural responses.

Table 4: Heterogeneity by Provincial Expansion Intensity

	Dependent variable: End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural × Post	0.031***	0.029***	0.033***
	(0.002)	(0.003)	(0.002)
Expansion × Post	0.004**	0.005**	0.002
	(0.002)	(0.002)	(0.002)
Rural × Post × Expansion	0.002	0.001	0.003
	(0.002)	(0.003)	(0.003)
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: *Expansion* is a standardised provincial measure of expansion intensity, defined as the average annual increase in the number of university seats between 1998 and 2006 at the provincial level, divided by the number of upper-secondary school graduates in 1998. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

9.2. Heterogeneity by Local University Access

Baseline higher-education availability is a policy-relevant proxy for pre-existing educational infrastructure and potential local exposure to tertiary opportunities. To assess whether the rural–urban divergence varies with this pre-reform capacity, Table 5 interacts $Rural \times Post$ with standardised provincial measures of higher-education density measured in 1998. Panel A uses the density of all higher-education institutions, and Panel B focuses on four-year universities.

Two patterns stand out. First, the $Post \times Rural$ coefficient is stable across panels and gender subsamples (approximately 0.03), indicating that the baseline divergence persists irrespective of how higher-education access is measured. Second, the triple interaction terms are consistently positive and precisely estimated: $Post \times Rural \times HE\ density$ is 0.010 in Panel A and 0.011 in Panel B (pooled), with similarly sized estimates for males and females. Given standardization, these coefficients imply that a one standard deviation higher baseline higher-education density is associated with an additional 1.0–1.1 percentage point increase in the post-reform rural–urban divergence in vocational high-school enrolment. Put differently, the divergence is more pronounced in provinces that had denser higher-education infrastructure before the reform, reinforcing the baseline finding while indicating meaningful variation in magnitude across pre-reform educational environments.

The lower-order $Post \times HE\ density$ terms are modest in magnitude. Where statistically significant (notably for females), they suggest small differences in overall post-reform changes across provinces with different baseline densities, but the primary heterogeneity in this table is concentrated in the differential rural–urban response captured by the triple interaction.

Table 5: Heterogeneity by Baseline Higher-Education Density

	Dependent variable: End at Vocational HS		
	(1) Pooled	(2) Male	(3)Female
Panel A: All higher-education institutions density			
Rural× Post	0.031***	0.029***	0.033***
	(0.002)	(0.003)	(0.002)
HE density × Post	-0.002	-0.000	-0.004***
	(0.001)	(0.002)	(0.002)
Post × Rural × HE density	0.010***	0.009***	0.010***
	(0.002)	(0.003)	(0.003)
Panel B: Four-year university density			
Post × Rural	0.031***	0.029***	0.033***
	(0.002)	(0.003)	(0.002)
Post × Four-year university density	-0.002	-0.000	-0.003**
	(0.001)	(0.002)	(0.001)
Post × Rural × Four-year university density	0.011***	0.011***	0.010***
	(0.002)	(0.003)	(0.003)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: Panel A uses the standardised provincial density of all higher-education institutions measured in 1998. Panel B uses the standardised provincial density of four-year universities measured in 1998. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

9.3. Heterogeneity by Regional Agricultural Dependence

Finally, we examine heterogeneity by pre-reform economic structure, measured by the agricultural share of GDP and grouped into terciles (low, middle, high). This dimension is relevant because it captures broad differences in local economic context that may correlate with the salience of post-reform educational opportunities. Table 6 reports a specification in which the *Post* × *Rural* effect is identified for the low-agriculture tercile and allowed to differ for the middle and high terciles via interaction terms.

The estimated rural–urban divergence is largest in low-agriculture provinces. In the pooled sample, *Post* × *Rural* in the low tercile is 0.041, while the additional terms for the middle and high terciles are negative (-0.008 and -0.019, respectively), implying smaller net effects as agricultural dependence rises. The attenuation is most pronounced in the high-agriculture tercile and is precisely estimated in the pooled sample. Gender-specific estimates follow a similar ordering, though differences are more sharply defined for females: the low-tercile effect is 0.049, with sizeable reductions in the middle (-0.014) and high (-0.031) terciles, yielding a notably smaller net divergence in more agriculture-dependent provinces. For males, the corresponding interaction terms are smaller and less precisely estimated, indicating weaker evidence of systematic differences across terciles.

Taken together, these results qualify the baseline finding in an interpretable way: the rural–urban divergence in vocational high-school enrolment is present across regions with different pre-reform economic structures, but it is less pronounced in more agriculture-dependent provinces, particularly among females. Importantly, the sign of the rural–urban divergence remains positive across terciles, and the main contrast is in magnitude rather than direction.

Table 6: Heterogeneity by Agricultural Dependence (GDP Share Terciles)

	Dependent variable: End at Vocational HS		
	(1) Pooled	(2) Male	(3)Female
Rural (Low agri share) × Post	0.041***	0.033***	0.049***
	(0.004)	(0.005)	(0.005)
Rural × Post × Middle agri share	-0.008	-0.002	-0.014**
	(0.006)	(0.008)	(0.007)
Rural × Post × High agri share	-0.019***	-0.006	-0.031***
	(0.005)	(0.007)	(0.006)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	343,886	175,277	168,534
County clusters	2,312	2,309	2,305

Notes: Agricultural dependence is measured by *Agriculture gdp share* and split into terciles (Low/Middle/High). The baseline category is the Low tercile . Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Across the three dimensions considered, the central empirical pattern—a post-reform divergence between rural and urban cohorts in vocational track enrolment—is robust. The divergence shows limited systematic variation with measured provincial expansion intensity, but it is larger in provinces with greater pre-reform higher-education density and smaller in more agriculture-dependent regions. These heterogeneity patterns are consistent with differences in pre-reform capacity and context shaping the magnitude of the reform-associated divergence, while leaving the qualitative baseline result intact.

10. Discussion

10.1. Main Findings

This paper documents an upstream response to China's 1999 higher-education expansion: cohorts exposed to the reform exhibit a clearer rural–urban divergence in upper-secondary terminal schooling, with rural individuals becoming more likely to end schooling at vocational high school relative to their urban counterparts. In magnitude, the post-expansion widening is on the order of a few percentage points in vocational high school completion, and the pattern appears persistent across exposed cohorts rather than confined to a single birth cohort. The divergence is present for both men and women, with broadly similar baseline magnitudes across gender.

Two forms of heterogeneity help characterise where the divergence is more pronounced. First, the rural–urban post-expansion shift toward vocational high school completion is larger in places with stronger pre-period educational capacity, as proxied by counties with higher baseline high-school attainment. Second, the divergence is also amplified in provinces with denser pre-reform higher-education infrastructure. In contrast, the divergence shows limited systematic steepening with measured provincial expansion intensity, suggesting that the core pattern is not confined to a small set of provinces that expanded most rapidly. Finally, across a range of placebo-style and differential-exposure exercises that allow the post-cohort rural–urban gap to vary with salient late-1990s shocks and policy initiatives, the central rural–urban divergence remains similar in sign and order of magnitude, indicating that it is not easily attributed to any single coincident event.

Taken together, these results underscore that a nationwide expansion at the tertiary level was accompanied by a detectable reconfiguration of upper-secondary endpoints, and that this adjustment was uneven across rural and urban populations and across pre-existing local educational environments.

10.2. Interpretations

The results suggest that tertiary expansions can reshape incentives and congestion upstream in a tracked education system. A plausible interpretation is a general-equilibrium reallocation: if the expansion raises the expected payoff to continued schooling, urban studentswho face less binding constraintsmay move out of vocational endpoints at higher rates, easing competition for vocational slots. This would make access to vocational schooling relatively less constrained for rural students, who remain more likely to sort into that track. The stronger divergence in areas with higher baseline capacity and denser higher-education infrastructure is consistent with upstream adjustments being mediated by local constraints: where capacity is higher, reallocation can operate more fully; where capacity is lower, responses are muted.

10.3. Implications

These findings speak to a broader debate on whether education expansions equalize opportunity or reallocate stratification across tracks. Even a nationwide tertiary reform can generate distributional changes earlier in the pipeline, shifting who occupies vocational tracks and how binding vocational constraints are. For policy design, the results highlight the value of coordinating reforms across levels: expanding tertiary capacity may change demand for upper-secondary pathways, and these responses may differ sharply between rural and urban populations and across local institutional environments.

10.4. Limitations

The analysis is based on terminal upper-secondary attainment and does not observe the full sequence of track entry, switching, or dropout decisions that generate final endpoints. Ruralurban status is measured at the census date and may not perfectly reflect location at the time schooling decisions were made, potentially attenuating measured gaps. While the stability of the results across multiple differential-exposure exercises reduces concern that a single contemporaneous shock drives the pattern, the paper does not isolate a unique micro-level mechanism. Finally, the analysis does not evaluate downstream outcomes such as labor-market returns or welfare, so it cannot speak to the efficiency or equity consequences of the compositional shifts documented here.

11. Conclusion

This paper documents an upstream consequence of China's 1999 higher-education expansion in a tracked secondary system. Using nationally representative census microdata that distinguish academic and vocational upper-secondary attainment, we show that cohorts exposed to the expansion experienced a persistent widening in the rural–urban gap in terminating schooling at vocational high school, on the order of a few percentage points. The divergence is larger in places with stronger pre-period educational capacity and denser pre-reform higher-education infrastructure, and smaller in more agriculture-dependent regions, while showing limited systematic steepening with measured provincial expansion intensity.

Taken together, the evidence suggests that tertiary expansions can reconfigure upper-secondary endpoints in ways that are not distribution-neutral across rural and urban populations, and that local educational environments shape the magnitude of these upstream adjustments. From a policy perspective, the results imply that higher-education massification may widen spatial inequality even when the reform is national in scope, by shifting who ends up in vocational upper-secondary schooling. This suggests that expansion policies may need complementary investments on the secondary margineither by easing rural constraints to remaining on the academic route (through capacity, information, or financial support), or by improving the quality and progression opportunities of vocational high schools so that a larger vocational endpoint does not translate into worse long-run prospects. More broadly, the findings highlight that local educational environments can mediate the incidence of national reforms and should be accounted for when evaluating distributional consequences.

An future direction is that the data speak most directly to terminal schooling outcomes rather than the full sequence of track entry and switching decisions. Future work that links administrative records on track assignment, exam participation, and school capacity to longer-run labour-market outcomes would help clarify the mechanisms and welfare implications of these upstream responses.

Data availability statement

The individual-level data used in this study come from the *2015 1% Population Sample Survey of China* microdata. Due to confidentiality restrictions and user agreements with the National Bureau of Statistics (NBS), the authors are not permitted to publicly release the raw microdata. Qualified researchers may apply for access through the NBS microdata application system.

Code availability statement

Replication code (Stata do-files and LaTeX source) is available from the corresponding author upon reasonable request.

Disclosure of interest

The authors report there are no competing interests to declare.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, we used OpenAIs ChatGPT to improve language clarity and readability. After using this tool, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research received no external funding.

References

Anderson, K. H., King, E. M., & Wang, Y. (2003). Market returns, transfers and demand for schooling in malaysia, 1976-89. *Journal of Development Studies*, *39*(3), 1–28.

Bennell, P. (1996). General versus vocational secondary education in developing countries: a review of the rates of return evidence. *The Journal of Development Studies*, *33*(2), 230–247.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, *3*, 1801–1863.

CCTV.com. (2002, July 27). *Rapid development of vocational education in china: Employment rate of secondary vocational school graduates reaches 78.4%.* CCTV.com, News Channel. Retrieved 2026-02-02, from https://tvguide.cctv.com/news/china/20020727/116.html (Published at 10:18 (Beijing Time))

Chen, D., Fu, N., & Pan, Y. (2019). Progress and challenges of upper secondary education in china. *World Bank Policy Research Working Paper Series*, *9042*. Retrieved from https://openknowledge.worldbank.org/handle/10986/31574

Dai, F., Cai, F., & Zhu, Y. (2022). Returns to higher education in china–evidence from the 1999 higher education expansion using a fuzzy regression discontinuity. *Applied Economics Letters*, *29*(6), 489–494.

Dai, L., & Martins, P. S. (2024). Does vocational education pay off in china? evidence from city-level education supply shocks. *Economic Modelling*, *140*, 106863.

Fan, W., & Ye, Y. (2021, June 21). *New china has achieved educational equity to the greatest extent.* China Youth Daily (newspaper), June 21, 2021, p. 03. Retrieved from https://zqb.cyol.com/html/2021-06/21/nw.D110000zgqnb_20210621_1-03.htm (Accessed 2026-02-01)

Filmer, D. (2007). If you build it, will they come? school availability and school enrolment in 21 poor countries. *The Journal of Development Studies*, *43*(5), 901–928.

Goldin, C., & Katz, L. F. (2018). The race between education and technology. In *Inequality in the 21st century* (pp. 49–54). Routledge.

Hao, L., Hu, A., & Lo, J. (2014). Two aspects of the rural-urban divide and educational stratification in china: A trajectory analysis. *Comparative Education Review*, *58*(3), 509–536.

Huang, B., Tani, M., Wei, Y., & Zhu, Y. (2022). Returns to education in china: Evidence from the great higher education expansion. *China Economic Review*, *74*, 101804.

Jia, R., Khanna, G., Li, H., & Xu, Y. (2025). The ripple effects of china's college expansion on american universities. *21st Century China Center Research Paper*(2025-04).

Krafft, C. (2018). Is school the best route to skills? returns to vocational school and vocational skills in egypt. *The Journal of Development Studies*, *54*(7), 1100–1120.

Kuczera, M., & Field, S. (2010). *Oecd reviews of vocational education and training: A learning for jobs review of china 2010: (chinese version).* Paris: OECD Publishing. doi: 10.1787/9789264113961-zh

Li, H., Loyalka, P., Rozelle, S., Wu, B., & Xie, J. (2015). Unequal access to college in china: How far have poor, rural students been left behind? *The China Quarterly*, *221*, 185–207.

Li, S., Whalley, J., & Xing, C. (2014). China's higher education expansion and unemployment of college graduates. *China Economic Review*, *30*, 567–582.

Liang, X. (2001). China: Challenges of secondary education. *Secondary Education Series, World Bank*, *22856*. Retrieved from https://documents.worldbank.org/en/publication/documents-reports/documentdetail/788771468771678465/china-challenges-of-secondary-education

Ou, D., & Hou, Y. (2019). Bigger pie, bigger slice? the impact of higher education expansion on educational opportunity in china. *Research in Higher Education*, *60*(3), 358–391.

Pyatt, G. (1966). *Human capital: A theoretical and empirical analysis, with special reference to education the residual factor and economic growth econometric models of education.* Oxford University Press Oxford, UK.

Ray, A., & Ruseski, J. E. (2026). Rural school investment and the rural-urban education gap: Evidence from india. *The Journal of Development Studies*, *62*(1), 68–82.

Standing Committee of the National People's Congress of the People's Republic of China. (1996). *Vocational education law of the people's republic of china.* English text accessed via the Ministry of Education of the People's Republic of China website. Retrieved from http://en.moe.gov.cn/documents/laws_policies/201506/t20150626_191390.html (Adopted May 15, 1996; effective September 1, 1996. Web version posted July 22, 2009. Accessed 2026-02-01.)

State Council of the People's Republic of China. (1999, January). *Notice of the state council forwarding the ministry of education's action plan for invigorating education toward the 21st century (guofa [1999] no. 4, jan 13, 1999).* Government document (reproduced online). Retrieved from https://www.shaanxi.gov.cn/zfxxgk/zfgb/1999/d3q_4444/200806/t20080626_1635800.html (Accessed 2026-02-01)

Wan, Y. (2006). Expansion of chinese higher education since 1998: Its causes and outcomes. *Asia Pacific Education Review*, *7*(1), 19–32.

Yeung, N. W.-J. J. (2013). *China's higher education expansion and social stratification.* Asia Research Institute, National University of Singapore.

Zhang, R. (2026). Higher education expansion and the rural–urban gap in secondary schooling and labor outcomes: Evidence from china's 1999 reform. *Journal of Development Economics*, 103721.

A. Schooling Stage at the 1999 Expansion

Table A.1: Schooling stage by birth cohort at the time of the 1999 higher-education expansion

Cohort	Age in 1999	Likely stage in 1998–1999 school yr.	Track-choice timing	Reform exposure
1979	19–20	Post-secondary / working	1995	No
1980	18–19	Senior(High) 3 *or* post-secondary	1996	First potentially affected (partial)
1981	17–18	Senior(High) 2–3	1997	Partial
1982	16–17	Senior(High) 1–2	1998	Partial
1983	15–16	Junior(Middle) 3 *or* Senior-entry	1999	Exposed (track choice after reform)
1984	14–15	Junior(Middle) 2–3	2000	Exposed (track choice after reform)

Notes: Ages are shown as ranges because cohort members can differ by up to one year at a given calendar date depending on birth month. 'Likely stage' therefore allows two adjacent grades/stages within the 1998–1999 school year, reflecting common variation in school entry age and grade progression (e.g., delayed entry or grade repetition). 'Track choice' refers to the upper-secondary transition at the end of junior secondary school. The 1980 cohort is highlighted as the first cohort for which the 1999 expansion could plausibly affect educational decisions (e.g., expectations and post-secondary plans), although most members had completed the upper-secondary track-choice decision prior to 1999; cohorts 1981–1983 are similarly classified as partially exposed. Cohorts 1984 and later reach the track-choice margin after 1999 and are classified as exposed.

B. Descriptive Trends by Pre-Period Educational Capacity

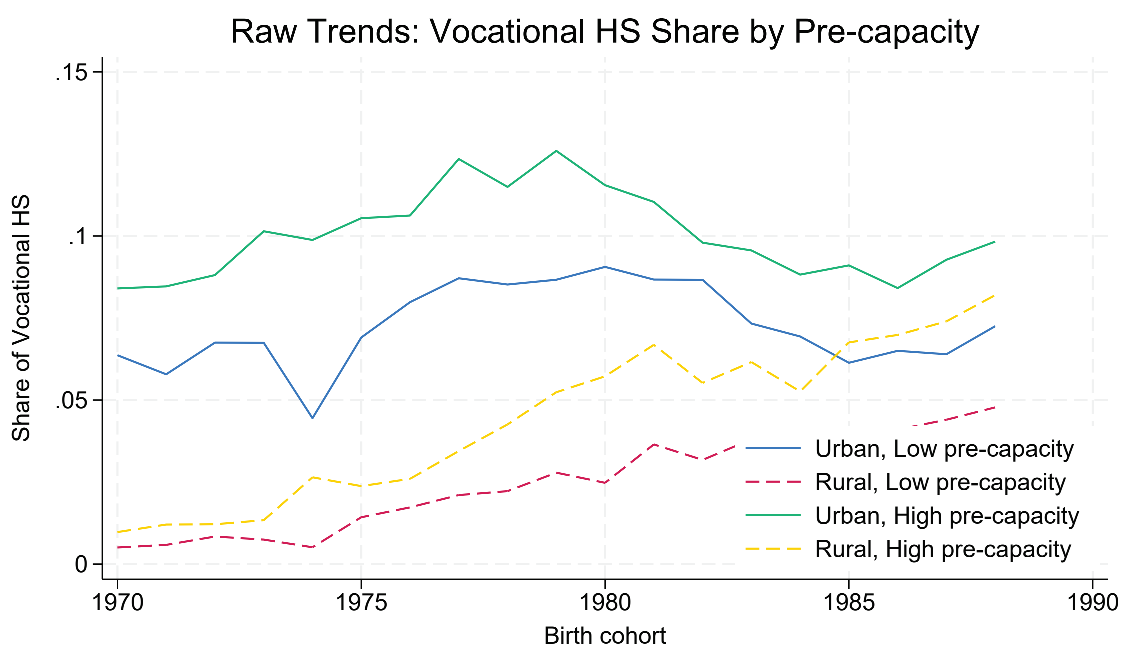

Figure B.1: Raw cohort trends in terminal vocational high-school attainment by county educational capacity

Notes: The figure plots cohort means of an indicator equal to one if the respondent's *highest completed* education is secondary vocational high school. 'Pre-period county educational capacity' is defined using the county-level share attaining any high school education among the 1974 birth cohort (above- vs. below-median). Lines are shown separately for rural and urban individuals within high- and low-capacity counties. These are unadjusted descriptive trends; the triple-difference estimates in Table 3 and the cohort-specific event-study results in Figure 3 additionally condition on controls and fixed effects and report county-clustered uncertainty.

C. Robustness Checks Tables and Figures

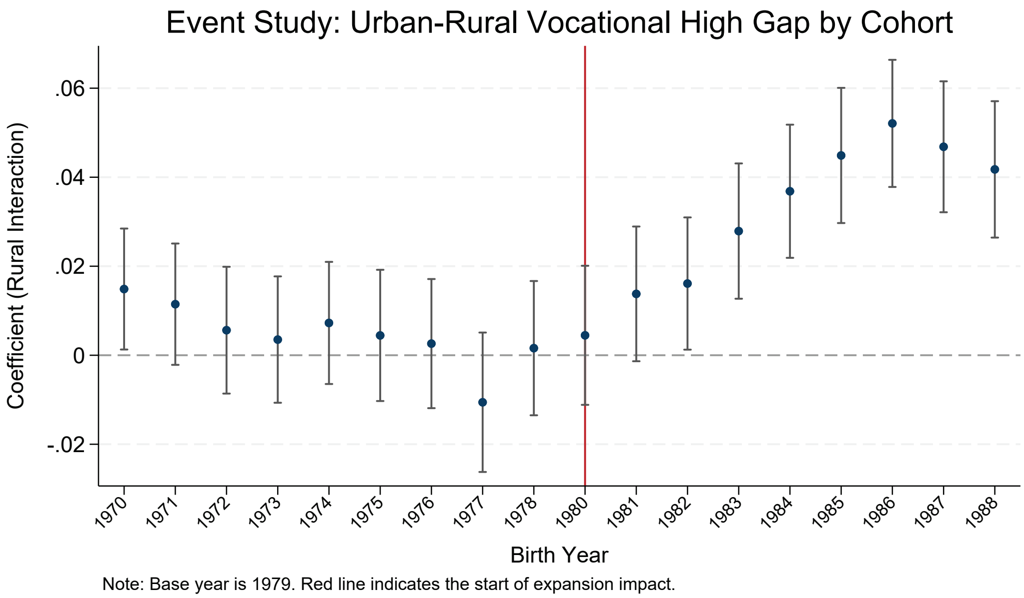

Figure C.2: Event Study: Hukou Non-movers (Rural–Urban Gap in Admission-Failure Share)

Notes: The figure plots cohort-specific estimates of the rural–urban gap in the share of vocational high school, normalised to the 1979 cohort. The sample is restricted to individuals whose hukou status did not change. Vertical bars indicate 95% confidence intervals. Standard errors are clustered at the county level. The working sample contains 226,133 individuals.

Table C.2: Robustness: Excluding Urban Districts

	End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural × Post	0.029***	0.027***	0.030***
	(0.002)	(0.003)	(0.002)
Controls	Yes	Yes	Yes
County and Cohort FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Observations	378,958	193,371	185,455
R^2	0.058	0.072	0.072
Clustering Level	County	County	County

Notes: The dependent variable is an indicator equal to one if an individuals highest completed level of schooling is vocational high school. This table re-estimates the baseline DiD specification after excluding prefecture-level city urban districts and retaining only non-district counties. Robust standard errors clustered at the county level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

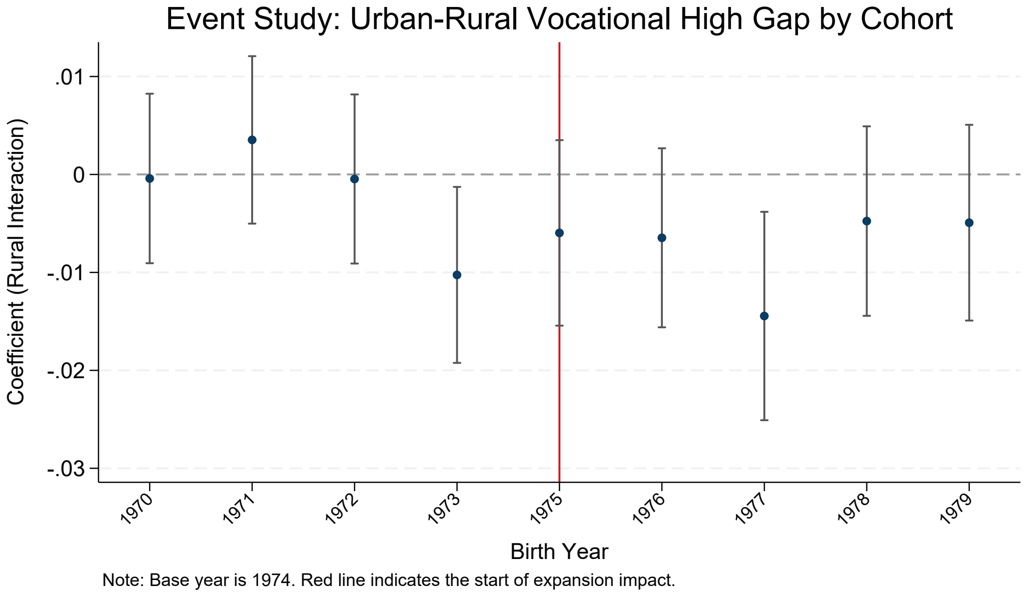

Figure C.3: Placebo Event Study: Pre-reform Cohorts (1970–1978)

Notes: The figure plots cohort-specific placebo estimates of the rural–urban gap in the share of vocational high school for cohorts born in 1970–1978, normalised to the omitted baseline cohort (1974). Vertical bars indicate 95% confidence intervals. Standard errors are clustered at the county level.

Table C.3: Cohort Window Sensitivity: Dropping 1979–1981 Cohorts

	End at Vocational HS		
	(1)Pooled	(2) Male	(3) Female
Post × Rural	0.036***	0.035***	0.037***
	(0.002)	(0.003)	(0.003)
Controls	Yes	Yes	Yes
County and Cohort FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Observations	348,273	177,837	170,301
Clustering Level	County	County	County

Notes: The dependent variable is an indicator equal to one if an individuals highest completed level of schooling is vocational high school. The sample excludes birth cohorts 1979–1981. *Post* is defined as cohorts born in or after 1975. Robust standard errors clustered at the county level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Alternative Clustering Levels

	End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Panel A: Clustered at Province			
Post × Rural	0.030***	0.028***	0.033***
	(0.004)	(0.004)	(0.005)
Controls	Yes	Yes	Yes
County and Cohort FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Observations	404,782	206,671	197,979
Clusters	31	31	31
Panel B: Clustered at City			
Post × Rural	0.030***	0.028***	0.033***
	(0.002)	(0.003)	(0.003)
Controls	Yes	Yes	Yes
County and Cohort FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Observations	404,782	206,671	197,979
Clusters	343	343	343

Notes: The dependent variable is an indicator equal to one if an individuals highest completed level of schooling is vocational high school. Robust standard errors are reported in parentheses and clustered at the level indicated in each panel.

D. Confounding Factors Tables

Table D.5: Accounting for the 1998 Asian Financial Crisis: Differential Exposure by Province Openness

	End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural × Post	0.036***	0.025***	0.049***
	(0.006)	(0.009)	(0.007)
High Region × Post	-0.000	-0.008	0.008
	(0.006)	(0.009)	(0.007)
Rural × Post × High Region	-0.007	0.003	-0.019***
	(0.006)	(0.009)	(0.007)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: This table examines whether the estimated post-cohort rural–urban gap in vocational high school attendance is confounded by differential exposure to the 1998 Asian Financial Crisis. *High Region* is an indicator for provinces with high crisis exposure (more open to external trade/FDI), and the omitted group is low-exposure provinces Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: Accounting for the 1998 Flood Shock: A County-Level Exposure Test

	End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural× Post	0.030***	0.028***	0.032***
	(0.002)	(0.003)	(0.002)
Flood × Post	-0.008	-0.006	-0.010
	(0.013)	(0.019)	(0.015)
Rural × Post × Flood	0.011	0.009	0.013
	(0.014)	(0.020)	(0.017)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: This table tests whether the estimated post-cohort rural–urban gap in vocational high school attendance is confounded by exposure to the 1998 flood shock. *Flood* is a county-level indicator equal to one for flood-affected counties in 1998. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.7: Accounting for the Nine-Year Compulsory Education Catch-up Campaign

	End at Vocational HS		
	(1) Full sample	(2) Male	(3) Female
Rural × Post	0.031***	0.029***	0.034***
	(0.002)	(0.003)	(0.003)
Campaign × Post	-0.005	-0.010	-0.001
	(0.006)	(0.009)	(0.007)
Rural× Post × Campaign	-0.005	-0.001	-0.010
	(0.006)	(0.010)	(0.008)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: This table tests whether the estimated post-cohort rural–urban gap in vocational high school attendance is confounded by exposure to Nine-Year Compulsory Education Catch-up Campaign. *Campaign* is a county-level indicator equal to one for campaign-affected counties in 1998. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.8: Accounting for the SOE Layoff Shock: Baseline Effects by Pre-Exposure SOE Employment Share

	End at Vocational HS		
	(1) Pooled	(2) Male	(3) Female
Rural × Post	0.031***	0.030***	0.033***
	(0.002)	(0.003)	(0.002)
SOE share × Post	-0.010***	-0.013***	-0.007***
	(0.002)	(0.002)	(0.002)
Rural × Post × SOE share	0.004**	0.006***	0.000
	(0.002)	(0.002)	(0.002)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: This table tests whether the estimated post-cohort rural–urban gap in vocational high school attendance is confounded by exposure to the late-1990s SOE layoff shock. 'SOE share' is the standardised provincial SOE employment share measured in 2000. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D.9: Accounting for WTO Accession: Continuous Exposure Based on Per-Capita Trade Volume

	End at Vocational HS		
	(1) Pooled	(2) Male	Female
Rural × Post	0.026***	0.027***	0.026***
	(0.004)	(0.005)	(0.004)
WTO exposure × Post	0.003**	0.004**	0.002
	(0.001)	(0.002)	(0.001)
Rural × Post × WTO exposure	0.002	0.001	0.003**
	(0.001)	(0.002)	(0.002)
County fixed effects	Yes	Yes	Yes
Birth-year fixed effects	Yes	Yes	Yes
County × Rural fixed effects	Yes	Yes	Yes
Observations	404,782	206,671	197,979
County clusters	2,841	2,826	2,823

Notes: This table tests whether the estimated post-cohort rural–urban gap in vocational high school attendance is confounded by differential exposure to China's WTO accession. *WTO exposure* is a continuous measure constructed from per-capita trade volume (higher values indicate greater exposure). Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E. Additional Tables and Figures

Table E.10: Provincial Expansion Intensity

Province	Expansion Intensity
Beijing	0.29
Tianjin	0.39
Hebei	0.21
Shanxi	0.22
Inner Mongolia	0.14
Liaoning	0.18
Jilin	0.17
Heilongjiang	0.20
Shanghai	0.24
Jiangsu	0.23
Zhejiang	0.23
Anhui	0.21
Fujian	0.27
Jiangxi	0.29
Shandong	0.20
Henan	0.22
Hubei	0.24
Hunan	0.22
Guangdong	0.24
Guangxi	0.19
Hainan	0.28
Chongqing	0.34
Sichuan	0.24
Guizhou	0.25
Yunnan	0.18
Tibet	0.12
Shaanxi	0.23
Gansu	0.16
Qinghai	0.22
Ningxia	0.26
Xinjiang	0.15

Notes: This table reports, at the provincial level, the average annual increase in the number of university seats between 1998 and 2006, divided by the total number of high school graduates in 1998.