From Seats to Status: China's 1999 Higher-Education Expansion and Urban-Rural Occupational Mobility

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

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

Abstract

China's 1999 higher-education expansion sharply increased college admissions through a centrally initiated supply shock. We study whether this expansion narrowed the rural–urban divide in occupational outcomes. Using the 2015 1% Population Sample Survey, we introduce an LLM-assisted occupational socioeconomic-status score that maps reported occupations into an interpretable status scale. We leverage cohort-based exposure to estimate impacts on rural–urban occupational convergence and education-driven occupational upgrading. We find that post-expansion rural cohorts gain 0.25 points in occupational status relative to urban cohorts, closing roughly 17% of the raw urban–rural gap. IV estimates provide evidence of a causal education channel, with each additional year of schooling raising occupational status by 0.21 points. Supply-side education policy shifts rural cohorts out of low-skill manual work and into skilled trades and basic services, while entry into professional and managerial positions remains unchanged.

Keywords: Higher-education expansion, Occupational mobility, Urban-rural disparity, Large language models, Text-based measurement

1. Introduction

Over the past half-century, higher education has expanded on a massive scale worldwide, drawing students from an ever-wider range of social and economic backgrounds into tertiary institutions.[1] Few policy changes expanded college access as quickly as China's 1999 reform, which sharply increased admissions through a centrally directed expansion. This paper asks whether that supply shock helped rural-origin cohorts move into higher-status occupations and narrowed China's rural–urban divide in career trajectories. Despite decades of rapid growth, the opportunity structure in China remains segmented: access to public services, local schooling options, and formal-sector jobs is still differentiated by urban versus rural origin through the household registration (*hukou*) system and related institutions (Chan & Buckingham, 2008). These institutional frictions shape not only wages at a point in time but entire job ladders—who enters stable, credential-intensive occupations and who remains concentrated in low-skill work.

Whether such a supply-driven expansion narrows or widens the rural–urban occupational gap is theoretically ambiguous. A standard human-capital channel predicts that relaxing capacity constraints raises degree attainment and thereby improves access to higher-status occupations (Becker, 1964). A credential/signaling channel implies that degrees may also re-rank workers when ability is imperfectly observed (Spence, 1973).

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

[1] As a magnitude benchmark, the global tertiary gross enrollment ratio increased from about 10% in the early 1970s to 44% in 2024. (World Bank, 2025)

At the same time, general-equilibrium forces can reshape hiring thresholds and the scarcity value of credentials as the graduate supply expands, generating credential inflation and heterogeneous impacts across rungs of the job ladder (Autor, Levy, & Murnane, 2003; Goldin & Katz, 2018). In a segmented labor market, these channels interact with durable urban advantage: rural youth may benefit disproportionately if the expansion relaxes a binding constraint, or they may fall behind if informational, resource, or institutional barriers prevent them from converting additional capacity into sustained occupational upgrading (Chan & Buckingham, 2008).

To empirically identify these effects, we exploit the 1999 higher education reform as a quasi-experiment. The expansion was a discrete, centrally initiated capacity shock: within a short window, authorities increased admissions slots and expanded institutional scale (Ministry of Education of the People's Republic of China, 2000). Aggregate statistics underscore the magnitude. New entrants into higher education increased by about 47% between 1998 and 1999 (National Bureau of Statistics of China, 2000). Because college entry is centrally mediated through the national College Entrance Exam and quota system, the expansion generates transparent cohort-based variation in exposure that is well suited to event-study diagnostics and quasi-experimental designs.

We focus on occupational status for two primary reasons. First, large harmonized population microdata typically record education and occupation for complete cohorts but rarely report earnings due to privacy constraints. Second, occupational position captures multidimensional socioeconomic status beyond mere wages. A common approach in stratification research is to map occupations into continuous status indices such as the International Socio-Economic Index (ISEI) (Duncan, 1961; Ganzeboom, De Graaf, & Treiman, 1992; Treiman, 1977). However, applying ISEI in China faces three distinct challenges. First, regarding contextual validity, the index is not calibrated for China' s specific labor-market structure. Second, regarding granularity, it assigns a single score to occupational titles, thereby masking critical heterogeneity in rank and sector. Third, regarding coverage, it results in substantial sample loss due to unmatched occupational codes.

We therefore develop a text-based occupational-status measure using a large language model (LLM). Specifically, we construct an LLM-assisted occupational socioeconomic-status score by first compiling and scoring a roster of common occupations under a transparent rubric, and then matching respondents' occupation descriptions to the scored roster to assign individual-level scores. This approach substantially improves granularity and coverage relative to standard indices, while offering a portable framework for other text-rich settings. More broadly, our work illustrates how modern 'text-as-data' tools can recover economically meaningful measures even when standard administrative codings are coarse or incomplete.(Gentzkow, Kelly, & Taddy, 2019; Grimmer & Stewart, 2013).

Empirically, we proceed in two steps. First, we document cohort dynamics in rural–urban occupational gaps using event-study difference-in-differences. Exploiting the cohort-based timing of exposure, we estimate cohort-by-rural interactions (normalized to a clean pre-reform baseline cohort) with county fixed effects and county-by-rural fixed effects, so identification comes from within-county comparisons of rural and urban individuals across adjacent cohorts. Standard event-study diagnostics provide a direct check of whether rural and urban cohorts exhibited differential pre-trends prior to the expansion.

Second, we provide causal evidence on the education channel using a shift-share instrumental variable strategy tailored to educational attainment. In shift-share designs, an aggregate shock is distributed across units using predetermined shares (Adão, Kolesár, & Morales, 2019; Bartik, 1991; Borusyak, Hull, & Jaravel, 2022, 2025; Goldsmith-Pinkham, Sorkin, & Swift, 2020). Because our treatment is years of schooling rather than an industry outcome, we construct county-level predicted exposure by interacting (i) province-level expansion intensity over 1998–2006 with (ii) a predetermined county-level measure of the population "at risk" of college entry: the county high-school completion rate among cohorts born in 1960–1965. We then interact this exposure with a post-reform cohort indicator to obtain a cohort-varying instrument. This design can be interpreted as a fuzzy difference-in-differences with an instrumented first stage (De Chaisemartin & d' Haultfoeuille, 2018; Hudson, Hull, & Liebersohn, 2017; Miyaji, 2024). Importantly, DID-IV identification requires two parallel-trends conditions: in the absence of the reform, high- and low-exposure counties must have followed similar trends not only in outcomes but also in the endogenous treatment. Following recent methodological guidance, we therefore assess pre-trends separately for the first stage (schooling) and the reduced form (occupational outcomes), and we report joint pre-trend tests for all pre-policy leads (Hudson et al., 2017; Miyaji, 2024). We also report weak-instrument-robust inference when instrument strength is limited (Andrews, Stock, & Sun, 2019).

Our findings reveal meaningful occupational upgrading and partial rural catch-up, but also suggest that the expansion did not fully erase structural segmentation. Our reduced-form event-study estimates show no systematic pre-trends: pre-reform coefficients are small and close to zero, while post-reform cohorts exhibit a clear upward shift in rural outcomes relative to urban cohorts. Consistent with this pattern, the DID summaries indicate that post-expansion rural cohorts gain about 0.25 points in the LLM-based occupation score relative to urban cohorts, with nearly identical effects for men and women. This magnitude is economically meaningful: 0.25 points corresponds to roughly 7% of the overall mean occupation score (3.72) and closes about 17% of the raw urban–rural gap (1.48 points). Complementary evidence using a low-skill indicator shows that the expansion reduces rural individuals' probability of low-skill employment by roughly 3.00 percentage points relative to urban individuals. A tier-level decomposition further suggests that the convergence operates primarily through exits from low-skill manual work into mid-tier employment, with little evidence of gains at or above the white-collar threshold.

Instrumental-variable estimates support a causal education channel. In rural areas, additional schooling translates into sizeable occupational upgrading: one extra year of education raises the occupation score by about 0.21 points, implying—under a simple linear extrapolation—that a four-year college degree lifts the score by roughly 0.85 points. Given the raw urban–rural gap of 1.48 points, this magnitude suggests that a college degree could close about 58% of the gap for those whose education increases because of the expansion. In urban areas, the implied occupational return is larger, though estimates should be interpreted with some caution given weaker first-stage strength and the relevance of weak-instrument concerns (Andrews et al., 2019). Together, the cohort dynamics and these magnitudes are consistent with a mechanism in which the capacity shock raised tertiary attainment for marginal cohorts and thereby facilitated entry into higher-status occupations, while persistent institutional and labor-market barriers still limit how fully education translates into convergence along the occupational ladder.

This paper contributes to three literatures. First, it adds new evidence on how large expansions in higher education reshape long-run labor-market outcomes, emphasizing occupations and job ladders rather than wages alone (Card, 1999; Goldin & Katz, 2018; Oreopoulos & Salvanes, 2011). Second, it speaks to research on urban–rural inequality and institutional segmentation in China by quantifying how a major national policy shock translated into differential career trajectories across origin groups (Chan & Buckingham, 2008). While prior work documents persistent rural disadvantages in access to elite education (H. Li, Loyalka, Rozelle, Wu, & Xie, 2015), we complement this literature by quantifying how a major national policy shock translated into downstream occupational upgrading and convergence in career trajectories. Third, it contributes a portable text-based measurement framework: an LLM mapping from narrative descriptions to a standardized metric, implemented here for occupational socioeconomic status but readily extendable to other text-coded attributes in microdata(Athey, Brunborg, Du, Kanodia, & Vafa, 2024; Gilardi, Alizadeh, & Kubli, 2023).

2. Institutional Background

The initiation of China's higher-education expansion in 1999 marked the onset of one of the largest capacity shocks in educational history. Motivated by a desire to stimulate domestic demand and to defer labor market entry for young cohorts, the State Council dramatically increased admission quotas. The magnitude is visible in aggregate statistics: new entrants into general universities rose from 1.08 million in 1998 to 1.60 million in 1999—a 47% increase in a single year—and reached 2.21 million by 2000 (National Bureau of Statistics of China, 2000). This expansion was explicitly a supply-side policy: the government expanded physical capacity and authorized universities to admit more students, creating a sharp discontinuity in the probability of college entry for cohorts turning 18 around 1999.

Access to this expanded capacity was mediated through China' s centralized National College Entrance Examination. The system is rigidly structured: students take the exam in their final year of high school, and admission depends strictly on achieving a score above province-specific cutoffs determined by available quotas. This centralized structure implies that the 1999 shock differentially affected specific birth cohorts. Individuals born around 1980–1981, who reached college age just as the expansion hit, faced a significantly more favorable admissions environment than those born just a few years earlier. Unlike gradual demand-driven expansions, this policy generated a discrete jump in opportunity determined by an individual's birth cohort and the timing of the reform.

This supply shock coincided with the fundamental marketization of China's labor allocation. The late 1990s saw the dismantling of the traditional "Iron Rice Bowl" through massive state-sector restructuring, which shifted labor demand toward the burgeoning private sector. By the time the first post-expansion cohort graduated in 2003, the administrative job assignment system had been replaced by a market-oriented two-way selection system. Consequently, the occupational outcomes we observe reflect market-clearing forces and private-sector growth rather than state planning.

Despite these market reforms, opportunities remained segmented by the Household Registration (hukou) system. Historically, hukou status strictly regulated internal migration and defined access to public services, creating a dual structure between urban and rural populations (Chan & Buckingham, 2008). For rural students, the education system represented one of the few legitimate channels to obtain urban status and escape agricultural employment. However, rural youth historically faced significant barriers, including underfunded primary schools and lower admission rates compared to their urban peers. In this context, the 1999 expansion potentially altered the rigorous competition for social mobility, determining whether the new capacity would allow rural outsiders to access credential-intensive, formal-sector occupations previously dominated by the urban elite.

3. Data

3.1. Data Source and Analytical Samples

Our primary data source is the microdata from the 2015 1% Population Sample Survey of China (hereafter, the "2015 1% Survey"), administered by the National Bureau of Statistics. The survey is designed to be nationally representative and covers roughly one percent of the population, comprising over one million individuals. It contains detailed information on educational attainment, demographics, household characteristics, and geography, and records both standardized occupation codes and respondents' verbatim occupation descriptions.

We focus on individuals born between 1970 and 1988 (ages 27–45 in 2015). This window brackets the cohorts plausibly exposed to the 1999 higher-education expansion at the upper-secondary–to–college margin, while ensuring that schooling is largely completed and that early labour-market outcomes are observable by 2015. Cohorts born in 1970–1978 had typically completed high school before the 1999 expansion and therefore serve as a pre-reform baseline; we use 1978 as the last pre-reform cohort to obtain a cleaner control group, as earlier cohorts include a nontrivial share of individuals with delayed school entry and grade repetition who may still be at the high-school-to-college margin around 1999 and hence partially exposed to the expansion. Cohorts born in 1979–1988 were still in secondary school or approaching the high-school-to-college transition and could adjust upper-secondary track choices and postsecondary plans. We avoid later cohorts to reduce contamination from subsequent reforms (the 2006–2007 abolition of rural compulsory-education fees) and to ensure sufficient time for labour-market outcomes to stabilize.[2]

Because outcome information in the census is reported only for relevant subpopulations—for example, occupational information is observed only for individuals who report work—we use the largest available sample for each outcome of interest within a common birth-cohort window. All analyses draw on individuals born between 1970 and 1988, ensuring that comparisons across outcomes are made for cohorts exposed to the same institutional environment around the 1999 higher-education expansion.

Educational outcomes are observed for the full analysis sample, while labour-market outcomes are measured among respondents with valid occupation information. In each case, sample restrictions are dictated by data availability rather than research design choices, and all variables are constructed consistently across subsamples. Unless otherwise stated, we apply the person weights provided by the survey to maintain national representativeness, and we harmonize variable definitions and coding rules across all specifications.

3.2. Rural Status and Background Covariates

The survey does not provide a clean, directly comparable measure of hukou type in the microdata extract we use. We therefore 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, making land-contract rights a tight proxy for rural status in our context. We report descriptive statistics for the main analysis sample, stratified by rural status and gender and separately for pre- and post-reform birth cohorts, in Table A.7.

3.3. Occupational Socioeconomic Status

To measure occupational socioeconomic status, we construct two complementary occupation-based outcomes: an LLM-generated occupation score based on text occupation descriptions, and an International Socio-Economic Index of Occupational Status (ISEI) score based on standardized occupation codes.

[2] 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 post-1999 cohort dynamics.

3.3.1. LLM-assisted roster-matched occupation score

We construct the occupation score in two stages. First, we compile a comprehensive roster of common occupation entries. We then use Claude (Anthropic) to assign each roster entry a 1–10 socioeconomic-status score based on work content, skill requirements, and social standing under a transparent rubric. Second, we map each respondent's raw free-text occupation description to the pre-scored roster. We achieve this primarily through exact string matching. For residual descriptions that lack an exact match, we employ a lightweight text-embedding method to compute cosine similarity, assigning the score of the closest roster occupation subject to a predefined similarity threshold.

We deliberately rely on respondents' original free-text answers rather than standardized occupation codes. The verbatim responses preserve the raw information reported by respondents, whereas standardized codes are assigned during or after data collection and may introduce coding errors. More importantly, coded occupation categories often mask substantial socioeconomic heterogeneity within the same category, compressing meaningful variation in the occupational hierarchy. Two examples illustrate the importance of retaining within-category detail. In maritime transport, a single category can include "crew member" (score 4), "second engineer" (score 5), and "captain" (score 7). In the judiciary, one category may encompass "court clerk" (score 5), "judge" (score 8), and "presiding judge" (score 9). Scoring the text descriptions allows us to recover within-category variation that would otherwise be lost. Appendix Table A.5 provides additional within-code examples from several large occupation codes, and Appendix Table A.6 quantifies the magnitude of this information gain: five-digit occupation codes explain about two-thirds of the variation in our text-based scores ($R^2 = 0.669$), leaving roughly one-third as within-code variation that would be mechanically discarded under a code-level lookup.

We adopt a parsimonious 1–10 integer scale, rather than a 0–100 continuous scale, because the socioeconomic ranking of many occupations is inherently coarse and cannot be meaningfully distinguished at very fine margins. For example, a nanny and a security guard both map naturally into the same broad stratum (score 3) in our rubric; forcing a strict 0–100 ordering would require arbitrary distinctions that are difficult to justify and may introduce additional noise and bias. In this sense, our approach favors what we call "fuzzy correctness" over "precise error".

In implementation, we compiled an occupational roster containing about than 23,000 unique occupation entries. These entries were scored using Claude (Anthropic) on a 1–10 socioeconomic status scale.[3] The model was instructed to evaluate each occupation based on actual work content, skill requirements, and social standing, and to output a single integer score for each entry. Respondents' occupation descriptions that were uninterpretable or could not be reliably matched to our roster were coded as missing. To visualize how occupational status differs by hukou, Figure 1 plots the distribution of the LLM occupation score by hukou status. Panel (a) shows the histogram of the score, and Panel (b) reports the corresponding cumulative distribution functions (CDFs). Figure A.6 further depicts unadjusted cohort trends in the mean occupation score for urban and rural respondents.

We claim that this LLM method is robust and able to reproduce. For transparency and reproducibility, we provide the full LLM prompt, scoring rubric, examples, and implementation details, along with some discussion in the Appendix A.1. Table 1 summarizes the 1–10 scale, provides reference occupations for each stratum, and reports the pooled distribution of the occupation score in our analysis sample. To assess reproducibility, we scored our occupational roster six times independently. Across all pairwise comparisons,

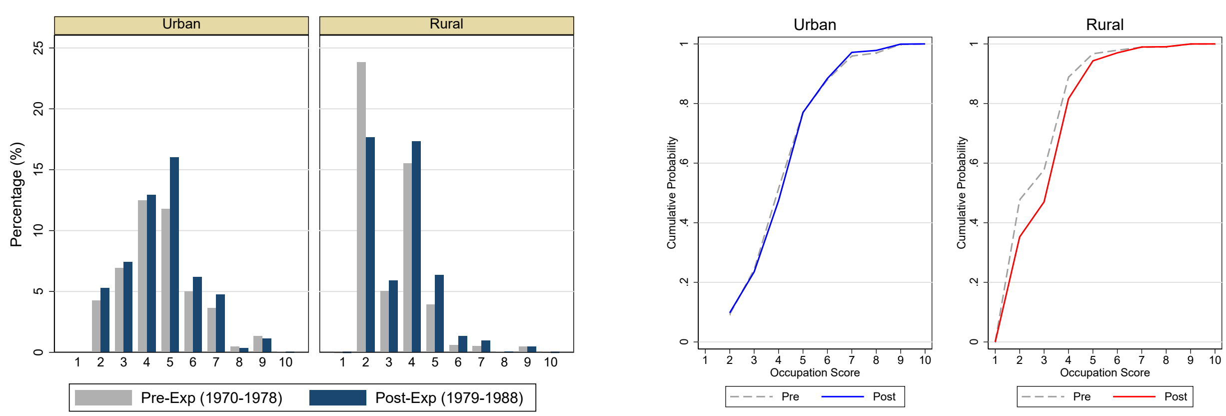
[3] Model version: `claude-opus-4-5-20251101`.

the resulting integer scores are identical for 92.0% of descriptions on average; allowing for ±1 point of tolerance, 96.8% of descriptions agree. The mean absolute deviation (MAD) across runs is 0.11 points. To further improve reproducibility, we employ a simple voting procedure: we query the model three times for each description and assign the modal score; when all three scores differ, we take the median. Comparing two such independently voted score sets, the exact agreement rate rises to 94.8% (±1: 98.0%), and MAD drops to 0.07 points.

Table 1: LLM-assisted roster-matched occupation score: Definition, Examples, and Distribution

Score	Socioeconomic Stratum	Reference Occupations	Freq.	Percent
1	Marginal survival	Scavenging	323	0.10
2	Low-skill manual labor	Agricultural worker; cleaner; porter; dishwasher	94,140	29.21
3	Basic service	Waiter; security guard; nanny; cashier	28,556	8.86
4	Trades/technical worker	Electrician; cook; driver; welder	99,505	30.88
5	Routine white-collar	Clerk; salesperson; cashier; receptionist	61,246	19.01
6	Junior professional	Primary school teacher; nurse; accountant	19,934	6.19
7	Mid-level professional	Secondary school teacher; physician; engineer	11,732	3.64
8	Senior professional	Associate professor; senior engineer; CPA	1,013	0.31
9	Executive/senior management	Professor; chief physician; general manager; factory director	3,192	0.99
10	Top leadership/businessman	Chairman; president; principal; party secretary	2,597	0.81
Total			**322,238**	**100.00**

Notes: This table describes the LLM-based 1–10 occupational socioeconomic score used in the paper. Columns "Freq." and "Percent" report the distribution in the pooled analysis sample. "Reference occupations" provide illustrative examples for each stratum and are not intended to be exhaustive.



(a) Distribution of Occupation Scores

(b) CDF of Occupation Scores

Figure 1: Changes in Occupation Score (LLM) Distribution by Hukou Status

3.3.2. ISEI occupational status score

As a benchmark measure of occupational socioeconomic status, we also use the International Socio-Economic Index of Occupational Status. The ISEI was developed by Ganzeboom, De Graaf, and Treiman (1992) to provide a comparable, cross-country measure of occupational stratification whenever standard occupation codes are available. The index summarizes an occupation's socio-economic position on a single continuous scale by linking occupations to the education and earnings typically associated with them. In our application, we use an ISEI crosswalk constructed by the China Family Panel Studies (CFPS) team, which maps occupation codes to ISEI scores while adapting the mapping to the Chinese occupational context.

4. Empirical Strategy

4.1. Reduced-Form DID and Event Study

This section presents reduced-form evidence on how occupational status varies with education and how the rural–urban gap evolves across birth cohorts around the 1999 higher-education expansion. We estimate two complementary specifications using the LLM-based occupation score as the outcome. The first uses the cross-sectional variation in educational attainment within county and cohort cells to characterize the education–occupation relationship. The second uses an event-study DID design to trace cohort-by-cohort changes in the rural–urban gap relative to a pre-reform baseline cohort.

First, we document the cross-sectional education gradient in occupational status while absorbing local and cohort confounders:

$$Y_{ict} \;=\; \beta\, Educ_{ict} \;+\; \gamma' X_{ict} \;+\; \alpha_c \;+\; \delta_t \;+\; \theta_{c\times Rural} \;+\; \varepsilon_{ict}, \tag{1}$$

where Y_{ict} is the occupational status outcome (the LLM-based occupation score), $Educ_{ict}$ denotes educational attainment, and X_{ict} includes individual controls (gender and ethnicity). The fixed effects α_c, δ_t, and $\theta_{c\times Rural}$ absorb time-invariant county characteristics, cohort-wide shocks, and persistent rural–urban differences within counties, respectively. Standard errors are clustered at the county level.

Second, we implement an event-study DID design to trace cohort-by-cohort changes in the rural–urban gap relative to a clean pre-reform baseline cohort. Let $D_t^k = \mathbb{1}\{t = k\}$ denote cohort indicators, and normalize the baseline cohort to 1978. We estimate:

$$Y_{ict} \;=\; \sum_{k\neq 1978} \beta_k \big(D_t^k \times Rural_i\big) \;+\; \gamma' X_{ict} \;+\; \alpha_c \;+\; \theta_{c\times Rural} \;+\; \varepsilon_{ict}. \tag{2}$$

This specification includes county fixed effects and county-by-rural fixed effects, so identification comes from within-county comparisons of rural and urban individuals across cohorts. The coefficients β_k capture the rural–urban differential in cohort k relative to the 1978 cohort. The pre-reform coefficients provide a direct check of stability in the rural–urban gap across cohorts, while the post-reform coefficients describe how that gap changes for cohorts plausibly exposed to the expansion during secondary school and the high-school-to-college transition.

4.2. Instrumental Variable Strategy

OLS estimates of the returns to education confound causal effects with selection into schooling. Individuals who attain more education may differ from those who do not in unobserved dimensions—such as ability, motivation, or family background—that independently affect occupational attainment. To isolate the causal effect of higher education on occupation scores, we exploit China's 1999 higher education expansion as a source of plausibly exogenous variation in college access. The expansion, which roughly doubled national tertiary admissions between 1998 and 2006, was a centrally initiated supply-side shock whose province-level intensity was largely determined by the pre-existing capacity and administrative priorities of provincial education authorities rather than by county-level labor market conditions (Ministry of Education of the People's Republic of China, 2000; National Bureau of Statistics of China, 2000).

We adopt a shift-share instrumental variable design in the spirit of Bartik (1991), though our construction differs from the canonical formulation in one important respect. In the standard Bartik approach, an aggregate shock is distributed across units using industry employment shares as exposure weights (Borusyak et al., 2022, 2025; Goldsmith-Pinkham et al., 2020). In our setting, the treatment is educational attainment rather than an industry-level outcome, so industry shares are not a natural decomposition. Instead, we construct county-level predicted exposure to the expansion by interacting a province-level shift with a county-level share that captures the local population at risk of upgrading to higher education.

Formally, define the instrument for county c in province p as

$$\text{Exposure}_c \;=\; \underbrace{\text{ExpProv}_p}_{\text{shift}} \times \underbrace{\text{Share}_c}_{\text{exposure weight}}, \tag{3}$$

where ExpProv_p measures the intensity of higher education expansion in province p over 1998–2006, calculated as the proportional increase in provincial tertiary admissions, and Share_c is the high school completion rate among the 1960–1965 birth cohorts in county c. The shift captures province-level variation in the magnitude of the supply shock, while the share captures the county-level stock of individuals whose educational margin was plausibly affected by the expansion. Counties with higher pre-reform high school completion rates had a larger pool of potential college entrants and were therefore more intensely treated by the province-wide increase in admissions slots.

The identifying variation arises from the interaction of two sources of cross-sectional heterogeneity with a cohort-based treatment assignment. Cohorts born in 1979 or later reached college-entry age during or after the expansion, while earlier cohorts did not. We interact Exposure_c with a post-reform cohort indicator to form the Bartik-style instrument $B_{ic} = \text{Exposure}_c \times \mathbf{1}[\text{birth year}_i \geq 1979]$, which varies across counties and birth cohorts. The analysis sample includes individuals born between 1974 and 1986, providing five pre-reform cohorts and eight post-reform cohorts.

We estimate the following system of equations separately for rural and urban subsamples. The second stage is

$$Y_{ic} \;=\; \beta\,\widehat{\text{Educ}}_{ic} \;+\; \mathbf{X}_{ic}'\gamma \;+\; \alpha_c \;+\; \delta_t \;+\; \varepsilon_{ic}, \tag{4}$$

where Y_{ic} is the LLM-based occupation score for individual i in county c born in year t, Educ_{ic} is years of education, $\mathbf{X}_{ic}$ is a vector of individual controls including gender and minority status, α_c denotes county fixed effects, and δ_t denotes birth-cohort fixed effects. The first stage is

$$\text{Educ}_{ic} \;=\; \pi\,B_{ic} \;+\; \mathbf{X}_{ic}'\phi \;+\; \alpha_c \;+\; \delta_t \;+\; \eta_{ic}. \tag{5}$$

County fixed effects absorb time-invariant differences in local economic structure, educational infrastructure, and geographic characteristics that may independently predict both schooling and occupational outcomes. Birth-cohort fixed effects absorb nationwide trends in educational attainment and labor market conditions common to all counties. The coefficient β in equation (4) is identified from within-county, across-cohort variation in years of education that is predicted by differential exposure to the higher education expansion. Standard errors are clustered at the province level to account for the fact that the shift component of the instrument varies at the province level, yielding 31 clusters.

Three features of this design merit discussion. First, identification follows the "exogenous shares" logic of Goldsmith-Pinkham et al. (2020): we treat the county-level high school completion shares as the source of identifying variation and require that, conditional on county and cohort fixed effects, counties with higher pre-reform high school shares would not have experienced differential changes in occupational attainment absent the expansion. The shares are constructed from cohorts born in 1960–1965, who completed their education well before the 1999 reform, ensuring they are predetermined with respect to the policy. We verify that these shares are orthogonal to province-level expansion intensity in Appendix A.20.

Second, the design can be interpreted as a fuzzy difference-in-differences, where the instrument generates differential first-stage effects across counties while the reduced form captures the corresponding differential impact on occupation scores (De Chaisemartin & d' Haultfoeuille, 2018; Hudson et al., 2017). The 2SLS estimator recovers the ratio of the reduced-form to the first-stage effect, which, under a monotonicity assumption requiring that the expansion weakly increased educational attainment for all compliers, identifies a local average treatment effect for individuals whose education was shifted by the reform.

Third, we estimate all specifications separately for rural and urban subsamples rather than pooling, because exposure to the expansion and the returns to education may differ systematically across these contexts. Rural areas had lower baseline college enrollment rates and fewer alternative pathways to high-skill occupations, so the expansion may have operated on a different margin of the educational attainment distribution. Separate estimation also allows us to assess instrument strength in each subsample independently, which is important given the well-known bias properties of 2SLS with weak instruments (Staiger & Stock, 1994; Stock & Yogo, 2002). Where the first-stage F-statistic—assessed via the Kleibergen and Paap (2006) robust Wald statistic—falls below conventional thresholds, we supplement standard 2SLS inference with Anderson–Rubin confidence intervals that are robust to weak identification (Finlay & Magnusson, 2009).

5. Result

5.1. Main Effects: Occupational Upgrading and Rural–Urban Convergence

Figure 2 presents cohort-based event-study estimates of the rural–urban occupational gap around the higher-education expansion. Each panel plots coefficients on cohort-by-rural interactions (normalized to zero for the 1978 cohort) with 95% confidence intervals. Across both occupational status measures, the pre-1979 coefficients are small and do not display a systematic trend, providing visual support for the identifying assumption that rural and urban cohorts would have followed similar trajectories absent the policy change. In contrast, the post-1979 cohorts exhibit a clear upward shift in the rural-relative-to-urban gap, consistent with occupational upgrading among rural cohorts exposed to the expansion.

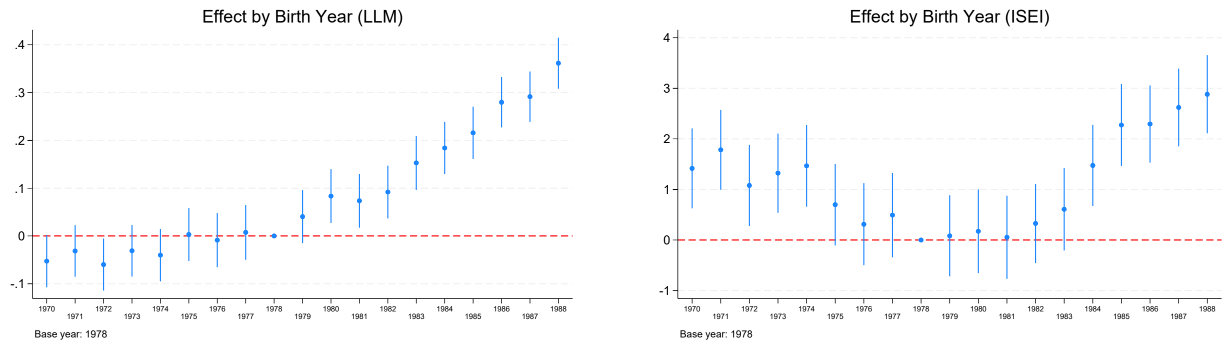
To summarize these cohort dynamics in a single estimand, Table 2 reports baseline difference-in-differences estimates of the change in the rural–urban occupational gap between post- and pre-expansion cohorts. Panel A shows that the post-expansion rural cohorts experience a statistically precise increase of 0.253 points in the LLM-based occupation score relative to their urban counterparts . The estimates are remarkably similar by gender (0.258 for men and 0.257 for women), indicating broadly comparable rural catch-up in this measure. Panel B provides complementary evidence using the conventional ISEI occupational status score: the point estimate is positive in the full sample and is more precisely estimated for men, while the estimate for women is smaller and imprecisely estimated.

As a complementary check, we examine whether the convergence in occupation scores is mirrored by a decline in the concentration of rural cohorts in low-skill employment. Unlike the LLM-based score, which captures fine-grained variation across ten tiers, the low-skill indicator is a coarse binary classification derived directly from occupation codes—sacrificing distributional detail but offering the advantage of transparency and immunity to any idiosyncrasy in the LLM scoring procedure. Appendix Figure A.8 plots event-study estimates of the rural–urban gap in low-skill employment: the pre-1979 coefficients show no differential trend, while post-1979 cohorts exhibit a clear and persistent downward shift. The corresponding DID estimate (Appendix Table A.13) indicates that the expansion reduced rural individuals' probability of holding a low-skill occupation by 3.1 percentage points relative to urban individuals, with a larger reduction for men (−4.0 p.p.) than for women (−2.3 p.p.). These patterns confirm that the occupational upgrading documented above operates not only through higher status scores but also through a reduced prevalence of bottom-tier employment among rural cohorts.

Table 2: Baseline DID Estimates of Rural–Urban Convergence in Occupational Status

	(1) Full sample	(2) Male	(3) Female
Panel A: Rural–Urban Difference-in-Differences (DV: LLM Occupation Score)			
Post × Rural	0.253***	0.258***	0.257***
	(0.012)	(0.016)	(0.017)
Dep. var. mean	3.722	3.845	3.564
Raw urban–rural gap	1.476	1.405	1.591
Observations	322,076	181,111	140,772
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	Yes	Yes	Yes
Panel B: Rural–Urban Difference-in-Differences (DV: ISEI Occupational Status Score)			
Post × Rural	0.352**	0.509***	0.163
	(0.141)	(0.182)	(0.196)
Dep. var. mean	48.681	47.422	50.011
Raw urban–rural gap	6.007	5.575	6.276
Observations	255,558	138,837	116,454
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	Yes	Yes	Yes

Notes: This table reports baseline difference-in-differences (DID) estimates of the change in the rural–urban occupational gap between post- and pre-expansion cohorts. Panel A uses the LLM-based occupation score as the dependent variable; Panel B uses the ISEI occupational status score. The coefficient on Post × Rural captures the differential change for rural individuals relative to urban individuals in post-expansion cohorts. All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. The main effects of *post* and *rural* are absorbed by the fixed effects and therefore omitted. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.



(a) LLM occupation score (event study). (b) ISEI occupation score (event study).

Figure 2: Event-study estimates of the rural–urban occupational gap by birth cohort
Notes: Each panel plots coefficients on cohort-by-rural interactions from the event-study specification, with pointwise 95% confidence intervals. The baseline cohort is 1978 (normalized to zero). Regressions include county fixed effects and county-by-rural fixed effects, controlling for gender and ethnicity; standard errors are clustered at the county level.

5.2. Granular Analysis of Occupational Mobility Across Tiers

To understand where in the occupational distribution the aggregate convergence originates, we decompose the effect into tier-specific changes. Figure 3 plots the 2SLS coefficient on Rural × Education for each occupation-score level (2–10), separately by gender; the corresponding reduced-form DID estimates and detailed coefficients are reported in Appendix Table A.11.

The figure reveals a striking pattern. For both genders, the dominant effect is a sharp reduction in the probability of holding a Score 2 (low-skill manual) occupation: the 2SLS coefficients are approximately -0.087 for men and -0.082 for women, and are the largest in absolute value across all tiers. The destination of this exit, however, differs markedly by gender. Men's upgrading is concentrated in Score 4 (skilled trades and technical work), where the coefficient spikes to about 0.079—roughly four times the magnitude of any other positive coefficient in the male profile. Women's gains are more evenly distributed across Scores 3 and 4 (basic services and skilled trades), with coefficients of 0.023 and 0.025 respectively. At Score 5 (routine white-collar), both profiles cross zero and remain close to it through the upper tiers, indicating that the education-driven convergence operates almost entirely through reallocation within the lower half of the occupational hierarchy.

The gender divergence is sharpest in the destination tiers. At Score 4 (skilled trades and technical work), the male coefficient (0.079) is more than three times the female coefficient (0.025), indicating that expansion-induced education gains channel rural men disproportionately into this tier. The pattern reverses at Score 3 (basic service work), where women's coefficient (0.023) is roughly twice the male estimate (0.011). This asymmetry suggests that the same causal increase in schooling sorts rural men and women into different segments of the mid-tier labor market: men leverage additional education to enter skilled manual and technical occupations, while women's upgrading is spread more evenly across basic services and skilled trades. From Score 5 onward, both profiles are close to zero, confirming that the education-driven convergence operates almost entirely within the lower half of the occupational hierarchy and does not extend to white-collar or professional tiers.

The tier-level decomposition clarifies the mechanism behind the aggregate convergence: the higher-education expansion enabled rural cohorts to exit the lowest-status occupations and move into skilled manual and technical work, but did not facilitate entry into white-collar or professional tiers. This "upgrading without breakthrough" pattern is consistent with a labor market in which college credentials open the door to mid-tier employment but where institutional barriers continue to gate access to higher-status occupations, such as hukou-based hiring preferences and urban social networks.

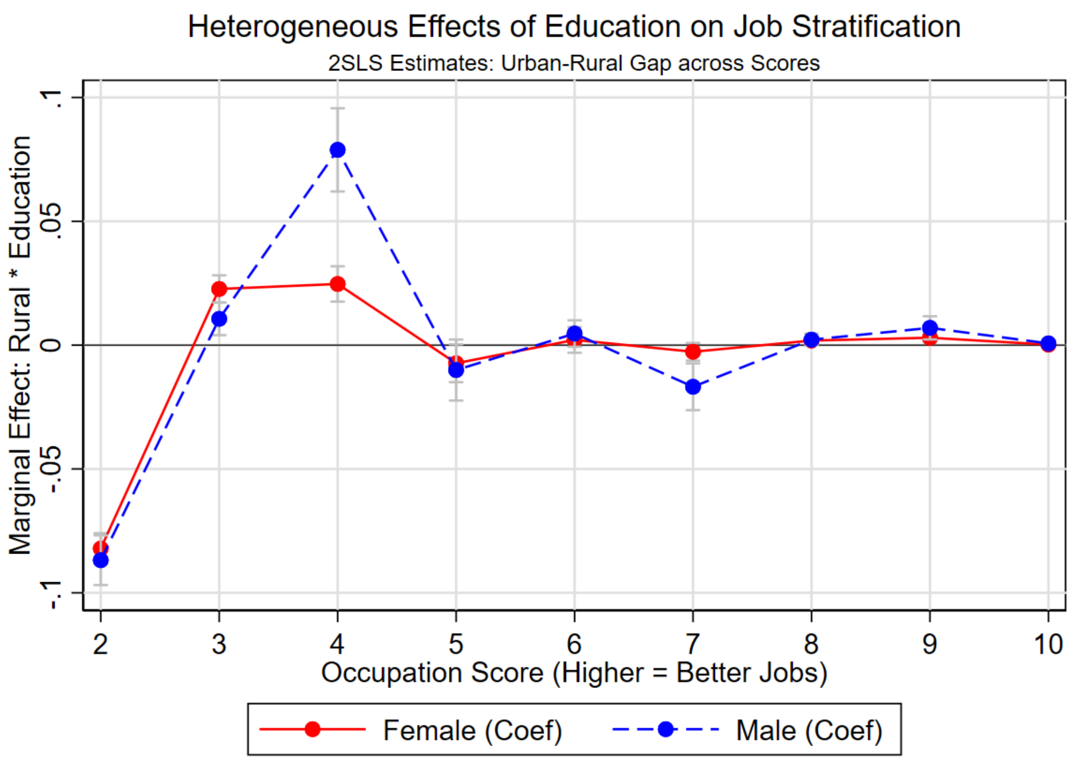


Figure 3: Heterogeneous Effects of Education on Job Stratification (2SLS Estimates)

Note: This figure plots the coefficients and 95% confidence intervals of the interaction term (*Rural* × *Education*) from 2SLS regressions across different occupation score levels. Panel A (red line) represents the female sample, and Panel B (blue line) represents the male sample.

5.3. Education as a Channel: OLS Benchmarks and IV Returns

The DID estimates above document rural–urban convergence in occupational status but do not identify the mechanism. A natural candidate is education: if the expansion increased college attainment among rural cohorts, and if college credentials improve occupational placement, then education is the proximate channel through which convergence operates. We test this logic in two steps—first documenting the descriptive association between schooling and the LLM-based occupation score, then isolating a causal effect using an instrumental variable strategy. Table 3 reports both sets of estimates in a unified framework, with each specification estimated under two clustering levels to gauge sensitivity to inference choices.

Because the LLM-based occupation score is a novel measure, we begin by benchmarking its units against years of schooling. Columns (1) and (2) of Table 3 report OLS associations between years of education and the occupation score, estimated separately by hukou status with the same fixed-effect structure used throughout the paper. Column (1) clusters standard errors at the province level (31 clusters) and Column (2) at the county level (approximately 2,700 clusters). The point estimates are identical across the two columns—only the standard errors differ. The goal is not causal identification but rather to establish that the score captures economically meaningful variation in occupational skill requirements and to provide a reference scale for interpreting magnitudes. The within-subsample gradient is 0.168 points per year in rural areas (Panel A) and 0.216 in urban areas (Panel B), both estimated very precisely under either clustering scheme. As a rough benchmark, our main DID estimate of 0.253 points for rural cohorts (Table 2) corresponds to roughly 1.5 years of rural schooling on this scale. These associations likely reflect both the causal effect of education and selection into schooling.

Columns (3)–(6) implement a shift-share instrumental variable strategy to isolate the causal component. Our instrument interacts province-level higher education expansion intensity (1998–2006) with predetermined county-level high school completion rates from the 1960–1965 birth cohorts. The first stage (Column 3) confirms instrument relevance: a one-unit increase in predicted exposure raises educational attainment by 3.2 years in rural areas and 1.8 years in urban areas. The reduced form (Column 4) shows that exposure directly raises occupation scores by 0.70 and 0.62 points, respectively. Scaling the reduced form by the first stage yields the 2SLS estimates in Columns (5) and (6), which present the same point estimates under province-level and county-level clustering. In rural areas, the 2SLS estimate is 0.214 points per year of education; in urban areas, 0.342 points per year.[4]

Because our instrument varies at the province–cohort level, province-level clustering is the more conservative choice, yielding larger standard errors and wider confidence intervals from only 31 clusters. County-level clustering exploits finer variation and produces tighter inference but may understate sampling uncertainty if within-province correlation in the instrument is important. We report both throughout to bracket the plausible range. Under province clustering, the Kleibergen–Paap F-statistics are 15.05 (rural) and 5.50 (urban); under county clustering, they rise to 39.74 and 10.64, respectively. The rural specification comfortably passes the Stock–Yogo 15% maximal IV size threshold under province clustering and the 10% threshold under county clustering. The urban specification falls short of conventional thresholds under province clustering but clears the 15% threshold under county clustering, warranting some caution in that panel.

To address remaining weak-instrument concerns we report Anderson–Rubin (AR) 95% confidence intervals, which remain valid regardless of instrument strength. All four AR confidence sets—rural and urban, under both clustering levels—exclude zero. In the rural subsample, the AR interval ranges from [0.03, 0.62] under province clustering to [0.05, 0.40] under county clustering; in the urban subsample, from [0.13, 0.53] to [0.14, 0.65]. The province-clustered intervals are wider, as expected given the small number of clusters, but the qualitative conclusion is robust: education causally raises occupation scores in both subsamples.

The comparison between OLS and IV estimates is informative about the nature of selection into schooling in each context. In rural areas, the 2SLS estimate (0.214) is close to the OLS estimate (0.168), consistent with minimal selection bias in a setting where college enrollment was historically constrained by supply-side factors—limited local high school quality and few university slots—rather than by unobserved ability differences among applicants. In urban areas, by contrast, the 2SLS estimate (0.342) exceeds OLS (0.216) by roughly 60 percent, a pattern consistent with positive returns heterogeneity among compliers. The expansion primarily shifted access for urban youth from less privileged backgrounds who had fewer alternative pathways to high-skill occupations; these marginal entrants may earn higher occupational returns from a college degree precisely because they lacked the family networks and institutional connections that substitute for formal

[4]Table A.12 in Appendix Appendix A.9 replicates this analysis using the International Socio-Economic Index (ISEI) as an alternative outcome. The rural 2SLS estimate is 2.573 ISEI points per year of education, broadly consistent with the pattern reported here. We prefer the LLM-based score for the main analysis because it captures occupation-specific skill content rather than mapping occupations to a predetermined prestige hierarchy; the ISEI results serve as a robustness check.

credentials in urban labor markets. Although all four AR confidence intervals exclude zero, the weaker first stage under province clustering for the urban subsample suggests that this interpretation, while supported by the data, should be held with somewhat greater uncertainty than the rural estimates.

To put the rural estimate in perspective, completing a four-year college degree raises occupation scores by roughly 0.85 points (0.214×4). Given a raw urban–rural gap of 1.476 points in the LLM-based occupation score (Table 2), a four-year degree would close approximately 58 percent of the gap for an individual complier. At the population level, however, the aggregate DID convergence of 0.253 points corresponds to roughly 17 percent of the raw gap—a meaningful but partial contribution, consistent with the tier-level evidence in Section 5.2 showing that education-driven upgrading is concentrated in the lower half of the occupational hierarchy.

A candid assessment of the IV strategy is warranted. Our primary contribution rests on the reduced-form DID estimates (Table 2 and Figure 2), which require only standard parallel-trends assumptions within a transparent county-by-rural fixed-effects structure and are supported by clean pre-trend diagnostics across all outcome measures. The shift-share IV serves a more circumscribed role: it disciplines the interpretation by confirming that education is a plausible causal channel and by providing an approximate magnitude for the return to schooling, but it necessarily rests on stronger assumptions. In particular, the county-level high school completion share ($Share_c$) may proxy for broader dimensions of local development—industrial structure, fiscal capacity, urbanization trajectory—that could independently shape occupational trends around 1999. We present a comprehensive battery of diagnostics in Section 7.3, including share orthogonality tests, first-stage and reduced-form pre-trend checks, specifications augmented with province-by-cohort fixed effects, leave-one-province-out sensitivity, alternative baseline-share windows, placebo first stages, exclusion-restriction tests, and monotonicity diagnostics. These exercises are broadly supportive, though we recognize that they have limited power given a limited number of pre-reform leads and 31 province-level clusters. The urban first-stage F-statistic falls below conventional thresholds under province clustering, warranting additional caution for that subsample. Readers who are skeptical of the IV assumptions can rely on the reduced-form DID results, which document the central empirical pattern—rural–urban convergence concentrated in the lower tiers of the occupational hierarchy—without requiring any exclusion restriction.

Table 3: The Effect of Higher Education on Occupation Scores: OLS and IV Estimates

	(1) OLS Occ. Score	(2) OLS Occ. Score	(3) First Stage Edu. Years	(4) Reduced Form Occ. Score	(5) 2SLS Occ. Score	(6) 2SLS Occ. Score
Panel A: Rural Subsample						
Years of education	0.168***	0.168***			0.214**	0.214**
	(0.004)	(0.002)			(0.104)	(0.086)
Bartik IV			3.245***	0.696**		
			(0.837)	(0.325)		
Observations	127,987	127,987	127,987	127,987	127,987	127,987
R^2	0.265	0.265	0.250	0.207		
KP Wald F					15.05	39.74
AR 95% CI					[0.03, 0.62]	[0.05, 0.40]
Panel B: Urban Subsample						
Years of education	0.216***	0.216***			0.342***	0.342***
	(0.004)	(0.002)			(0.068)	(0.108)
Bartik IV			1.806**	0.617*		
			(0.770)	(0.318)		
Observations	77,888	77,888	77,888	77,888	77,888	77,888
R^2	0.310	0.310	0.273	0.158		
KP Wald F					5.50	10.64
AR 95% CI					[0.13, 0.53]	[0.14, 0.65]
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	Province	County	Province	Province	Province	County

Notes: The dependent variable is indicated in the third header row: "Occ. Score" denotes the LLM-based occupation score; "Edu. Years" denotes years of education. Columns (1) and (5) report OLS and 2SLS estimates with standard errors clustered at the province level; Columns (2) and (6) repeat the same specifications with county-level clustering. Columns (3) and (4) report, respectively, the first-stage and reduced-form estimates with province-level clustering. The instrument is constructed as $Z_{ct} = ExpProv_p \times Share_c \times Post_t$, where $ExpProv_p$ measures province-level higher education expansion intensity, $Share_c$ is the county-level high school completion rate among the 1960–1965 birth cohorts, and $Post_t$ indicates cohorts born in 1979 or later. Individual controls include gender and minority status. The Kleibergen–Paap (KP) Wald F-statistic (Kleibergen & Paap, 2006) tests instrument strength; the Stock–Yogo critical values for one endogenous regressor are 16.38 (10% maximal IV size) and 8.96 (15%) (Stock & Yogo, 2002). Anderson–Rubin (AR) 95% confidence intervals, which are robust to weak instruments, are reported for both clustering levels in Columns (5) and (6) and are computed using the `weakiv` package (Finlay & Magnusson, 2009) on residualized data (after absorbing county and birth-cohort fixed effects). All four AR confidence sets exclude zero, confirming a positive causal effect of education on occupation scores under both clustering schemes. R^2 is omitted for the 2SLS columns because it lacks a standard interpretation under IV estimation. Sample: birth cohorts 1974–1986. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6. Heterogeneity Analysis

6.1. Heterogeneity by Agricultural Dependence

A natural question is whether the post-expansion rural–urban convergence varies with local economic structure. If occupational upgrading requires a non-agricultural labor market to absorb college-educated workers, convergence should be weaker in areas where agriculture dominates local output and high-skill job opportunities are scarce. To test this, we classify prefecture-level cities into terciles of pre-period agricultural output share and interact the baseline DID estimand with tercile indicators, so that Post $\times$ Rural captures the convergence effect in the least agricultural tercile (the omitted group) and the triple interactions measure departures from that baseline.

Table A.14 reports the results. The base effect in the low-agricultural tercile is positive and significant (0.176), confirming that convergence is present even in the most industrialized local economies. Strikingly, the interaction for the middle tercile is large and precisely estimated (0.175), implying that the total convergence effect in moderately agricultural areas is roughly twice the magnitude of the base effect. The high-agricultural tercile also exhibits additional convergence in the pooled sample (0.120), though the pattern differs by gender: the triple interaction remains large and significant for men (0.169) but is small and statistically insignificant for women (0.064).

These results suggest that the expansion-driven convergence is not confined to already-industrialized regions. Rather, the largest gains accrue in areas with moderate agricultural dependence—localities that retain a sizable rural population at risk of upgrading but also possess enough non-farm employment to absorb newly credentialed workers into higher-status occupations. The attenuation for women in the most agricultural tercile is consistent with the gender asymmetry documented in Section 5.1: women's occupational upgrading may be more sensitive to the availability of local service-sector and white-collar employment, which is thinnest in heavily agricultural areas.

6.2. The Evolution of the Urban-Rural Gap: Distribution and Returns

To rigorously examine the evolution of the urban-rural gap, we adopt a two-step approach. First, we analyze the distributional shifts in educational attainment to understand the changing composition of the workforce. Second, we employ a fixed-effects model to estimate how the returns to these educational credentials have evolved within local labor markets.

Descriptive Evidence: Asymmetric Expansion. Before estimating the returns to education, it is crucial to account for the "selection" into higher education. Figure 4 illustrates the distribution of educational attainment by hukou status across the pre- and post-expansion cohorts.

The figure reveals a striking asymmetry. In urban areas, the expansion effectively universalized higher education, with the college share approaching 50%. Conversely, in rural areas, while the college share increased, it remains a small fraction of the population (approximately 10%), with a significant "bottleneck" visible at the high school level.

This implies that while the urban college cohort has been heavily "diluted" (absorbing the median student), the rural college cohort remains highly positively selected—representing the academic elite of the rural population. This distributional context is critical for interpreting the regression results: if "negative selection" (lower ability students entering college) were the primary driver of falling returns, we would expect a sharper decline for the rapidly diluting urban group than for the selectively elite rural group.

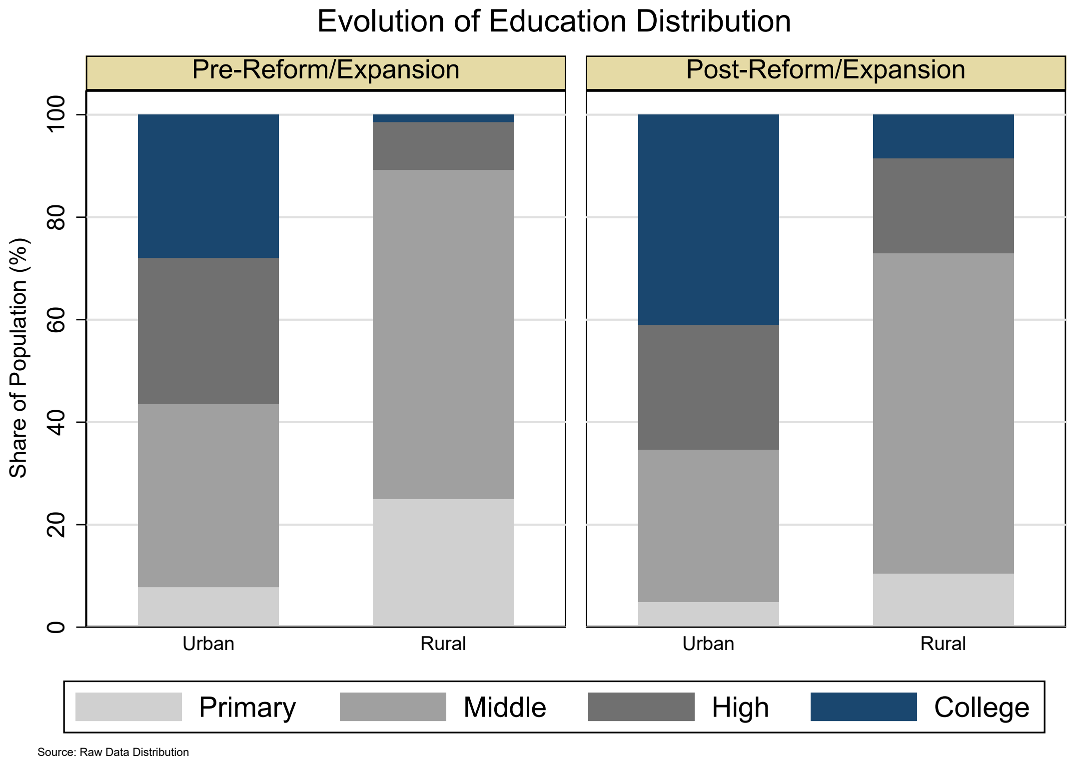


Figure 4: The Distribution of Educational Attainment: Urban vs. Rural (Pre- and Post-Expansion)

Note: This figure plots the share of each educational level within rural and urban populations based on birth cohorts. ”Pre” refers to cohorts born before 1979; ”Post” refers to cohorts born in 1979 or later.

Fixed-Effects Estimates. Next, to quantify the changing urban-rural gap, we specify the following fixed-effects equation:

$$Y_{ict} = \alpha + \beta_1 Rural_i + \sum_j \gamma_j Edu_{ij} + \sum_j \delta_j (Rural_i \times Edu_{ij}) + \lambda_c + \phi_t + \epsilon_{ict} \tag{6}$$

where Y_{ict} denotes the occupation score, and λ_c and ϕ_t represent county and birth cohort fixed effects. The coefficient δ_j is of central interest, capturing the *additional* gap (or premium) specific to education level j. Table 4 presents the results.

Narrowing Baseline Gap. The results indicate a significant reduction in the baseline institutional barrier. For the reference group (primary education), the rural penalty decreased from -0.434 to -0.262 (Column 1 vs. 2). This suggests that for the least educated workforce, market liberalization and the rise of rural industries have successfully compressed the urban-rural divide.

The Disappearance of the ”Equalizer” Effect. The most notable structural shift occurs at the top of the educational hierarchy. In the pre-reform period, the interaction term for *College* $\times$ *Rural* is significantly positive. This confirms that in the elite education era, a college degree served as a powerful ”equalizer,” allowing rural graduates to significantly offset their institutional disadvantage.

However, in the post-reform period, this interaction term becomes statistically insignificant (-0.041). This result is particularly stark when viewed in light of Figure 4. Despite rural college graduates remaining a highly selected elite (the top 10% of their cohort) compared to the mass-educated urban peers, they no longer enjoy a relative advantage in closing the gap. The data suggests a transition from an "Equalizer Model"—where high human capital could overcome hukou barriers—to a "Stratification Model," where the protective value of a degree has vanished for rural youth, leaving them structurally exposed to the same penalties as the general rural population.

Table 4: Urban-Rural Gap Analysis with County and Cohort Fixed Effects

	(1)	(2)
Dependent Variable: Occ. Score	Pre-Reform	Post-Reform
Base Gap (Primary Reference)		
Rural (Main Effect)	-0.434***	-0.262***
	(0.027)	(0.028)
Interaction Terms (Gap Change)		
Middle School × Rural	-0.273***	-0.120***
	(0.029)	(0.030)
High School × Rural	-0.243***	-0.219***
	(0.035)	(0.034)
College × Rural	0.125**	-0.041
	(0.053)	(0.034)
Observations	156,147	166,068
County FE	Yes	Yes
Cohort FE	Yes	Yes
County × Rural FE	Yes	Yes
Clusters (county)	Yes	Yes

Note: Standard errors clustered at the county level in parentheses.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

7. Robustness and Validity

This section summarizes a comprehensive battery of robustness and validity checks organized around three concerns: sensitivity of the reduced-form DID estimates to sample and specification choices, confounding from concurrent macroeconomic and policy shocks, and validity of the identifying assumptions underlying the shift-share instrumental variable design.

7.1. Robustness Checks of the Reduced-Form Estimates

A first concern is that the baseline DID estimates may be sensitive to sample composition, measurement choices, or inference procedures. We address this through seven robustness checks. Because the ISEI score is missing for a subset of observations, ISEI-based regressions use a smaller sample; re-estimating the LLM specification on the common sample with non-missing ISEI yields an essentially identical event-study profile and point estimate, ruling out compositional differences from sample availability (Appendix Figure A.9). To address the possibility that hukou reclassification or cross-county migration alters the composition of rural and urban samples across cohorts, we restrict to non-movers; the main patterns are unchanged (Appendix Figure A.10). Because municipal districts (*shixiaqu*) and direct-controlled municipalities may differ systematically in labor-market structure, we drop these units and confirm that the education–occupation gradient is virtually unchanged (Appendix Table A.15).

We also probe the cohort-window definition. A placebo event study on pre-reform cohorts (1970–1978) shows no systematic effects, supporting the absence of differential pre-trends (Appendix Figure A.11). Excluding the border cohorts born in 1978–1980—whose exposure to the expansion is ambiguous—yields a slightly larger but statistically indistinguishable point estimate of 0.286 (Appendix Table A.16). Finally, because the expansion-intensity shifter varies at the province level, we re-estimate the 2SLS specifications with province-clustered standard errors; the first stage remains strong and the second-stage estimates are stable in sign and magnitude (Appendix Table A.17). Across all exercises, the estimated post-expansion rural–urban convergence is robust.

7.2. Potential Confounders and Concurrent Shocks

Several major shocks and policy changes occurred in the late 1990s and early 2000s. A concern is that these events may have affected local labor markets or schooling environments in ways that differ between rural and urban residents, and that the affected places may systematically differ in pre-existing characteristics that also matter for educational supply and occupational outcomes. If so, the estimated post-1999 rural–urban divergence could partly reflect these concurrent shocks rather than the higher-education expansion.

To probe this concern, we implement a common robustness strategy across shocks. For each event, we define an *exposure* measure using characteristics determined before the event (or a predetermined indicator of whether a location was affected). We then allow the post-1999 rural–urban differential to vary with exposure, so that any confounding driven by differential exposure to the shock is absorbed by the interaction terms. Across all shocks examined, the post-1999 rural–urban differential remains positive and similar in magnitude to the baseline.

Asian Financial Crisis (1998). The 1998 Asian Financial Crisis could have altered local employment opportunities, especially in places with larger financial-sector activity. We proxy crisis exposure using the pre-crisis financial-industry employment share (high vs. low exposure) and estimate a triple-difference specification that allows the post-1999 rural–urban differential to vary with crisis exposure. Table A.18 shows that the post-1999 rural–urban differential remains positive and statistically significant. The additional differential by crisis exposure is small and statistically insignificant in the pooled and male samples; for females it is negative and statistically significant, implying a smaller—but still positive—post-1999 rural–urban differential in high-exposure regions.

Large flood (1998). The 1998 flood may have caused localized disruptions to schooling conditions and economic activity in affected counties. We classify counties by whether they were flood-affected and allow the post-1999 rural–urban differential to vary with flood exposure. Table A.19 shows that the post-1999 rural–urban differential is very similar to the baseline estimate, and the additional differential associated with flood exposure is statistically indistinguishable from zero.

Nine-year compulsory-education catch-up campaign. A separate concern is that the compulsory-education catch-up campaign may have been targeted to particular counties and could have changed schooling trajectories around the same period. Using a predetermined indicator for counties covered by the campaign, we test whether the post-1999 rural–urban differential differs between campaign and non-campaign counties. Table A.20 shows that the post-1999 rural–urban differential remains positive and highly significant. The additional differential is positive and statistically significant, indicating that the post-1999 rural–urban differential is larger in campaign counties; importantly, allowing for this heterogeneity does not overturn the main result.

State-owned enterprise (SOE) restructuring. SOE restructuring in the late 1990s could have generated large labor-demand shocks, particularly in provinces with high SOE employment prior to restructuring. We proxy exposure using the pre-restructuring SOE employment share (high vs. low exposure) and allow the post-1999 rural–urban differential to vary with SOE exposure. Table A.21 indicates that the post-1999 rural–urban differential remains positive and statistically significant. The additional differential by SOE exposure is close to zero and statistically insignificant in the pooled and male samples; for females it is negative and statistically significant, implying a smaller—but still positive—post-1999 rural–urban differential in high-SOE provinces.

China trade shock (WTO-related exposure). Finally, rising trade exposure around WTO accession may have reshaped labor demand differently across provinces. We use a pre-period measure of WTO-related exposure (above- vs. below-median) and test whether the post-1999 rural–urban differential varies with trade exposure. Table A.22 shows that the post-1999 rural–urban differential remains positive and precisely estimated, while the additional differential by WTO exposure is small and statistically insignificant. Overall, differential exposure to the China trade shock is unlikely to account for the estimated post-1999 rural–urban pattern.

7.3. Validity of the DID-IV Design

A third set of checks targets the identifying assumptions of the shift-share instrumental variable strategy: instrument relevance, the parallel-trends and no-anticipation assumptions, the exclusion restriction, and monotonicity.

Parallel trends and no anticipation. Event-study estimates for both the first stage and reduced form show no differential trends between high- and low-exposure counties in the pre-reform period: joint F-tests cannot reject that all pre-reform coefficients equal zero ($p > 0.10$ in both rural and urban samples), validating the parallel-trends assumption for both education and occupation scores (Appendix Figure A.12; Appendix Table A.24, Panel D). Placebo first-stage regressions using only pre-policy cohorts (1974–1978) with artificially early cutoffs yield small and mostly insignificant coefficients, inconsistent with anticipation effects (Appendix Table A.24, Panel A). The county-level high school shares used to construct predicted exposure are orthogonal to province expansion intensity ($p > 0.38$; Panel C), confirming they are predetermined.

Exclusion restriction. The exclusion restriction requires that the instrument affects occupation scores only through education. Three tests support this assumption (Appendix Table A.25). Adding province-by-cohort fixed effects—which absorb all province-level time-varying confounders—leaves the rural 2SLS estimate broadly consistent with the baseline (0.146 vs. 0.214), though precision declines as expected given the more demanding fixed-effect structure. Restricting to non-migrants yields estimates nearly identical to the baseline (rural: 0.215; urban: 0.344), ruling out spatial sorting as an alternative channel. The instrument does not predict cross-county migration in either subsample, confirming that migration is not a pathway through which the instrument operates independently of education.

Monotonicity. First-stage estimates are uniformly positive across pre-specified demographic subgroups—by gender, hukou status, and their interaction—consistent with the absence of systematic defiers (Appendix Table A.23).

Sensitivity to influential observations and construction choices. A leave-one-province-out jackknife confirms that no single provincial shock drives the results: the 31 rural LOPO estimates range from 0.168 to 0.261 (SD = 0.021) and the urban estimates from 0.315 to 0.371 (SD = 0.013), all tightly clustered around the respective baselines (Figure A.13). Re-computing the county exposure share using 55 alternative baseline cohort windows (start years 1960–1972; widths 2–6 cohorts) shows that the 2SLS estimate, instrument strength, and first-stage pre-trend diagnostics are insensitive to this construction choice (Appendix Figures A.14–A.16).

8. Discussion

The empirical results point to a pattern we characterize as "upgrading without breakthrough." We interpret this pattern by mapping each key finding to the mechanisms it supports or rules out.

Magnitude. The baseline DID estimate of 0.25 points corresponds to roughly 17% of the raw urban–rural gap in occupational status (1.476 points on the LLM scale), or about 1.5 years of schooling on the within-rural education gradient. This is an intent-to-treat effect averaged over all rural individuals, including the majority who did not enter college. The IV estimates isolate the complier-specific return: each additional year of schooling raises occupational status by 0.21 points. Because this estimate applies to compliers, it should be interpreted as an upper bound on the effect of college completion for the marginal rural entrant, not as the average effect across all rural youth. Both magnitudes are economically meaningful but clearly partial, consistent with the tier-level evidence showing that education-driven convergence is concentrated in the lower half of the occupational hierarchy.

Distinguishing mechanisms. Three channels could link the expansion to rural–urban convergence: a direct human-capital channel, a credential-inflation channel, and a persistent institutional-barrier channel.

The IV estimates support the human-capital channel: education causally raises occupational status for rural compliers. However, the tier-level decomposition reveals that the marginal effects are concentrated entirely at Scores 2–4 and are indistinguishable from zero at Score 5 and above. This rules out a simple story in which more schooling translates uniformly into better jobs across the occupational ladder. Instead, the evidence is consistent with institutional barriers—hukou-based hiring preferences, urban professional networks, and formal-sector credentialing requirements—continuing to gate access to white-collar and professional tiers independently of educational credentials.

The credential-inflation channel is supported by our subgroup estimates. In the pre-expansion period, a college degree served as a powerful equalizer: rural college graduates enjoyed a significantly smaller rural penalty than their non-college peers, suggesting that higher education could partially offset institutional disadvantage. In the post-expansion period, this equalizing advantage disappears, even though rural college graduates remain highly positively selected (roughly the top 10% of their cohort). This pattern is consistent with a general-equilibrium adjustment in which the scarcity value of credentials erodes as urban cohorts move toward mass college enrollment.

Gender asymmetry. The same causal increase in schooling channels rural men disproportionately into skilled trades (Score 4), whereas women's gains are spread more evenly across basic services and skilled trades (Scores 3–4). This divergence likely reflects gendered occupational sorting within the non-agricultural labor market rather than differential educational gains.

Policy implications. Supply-side education expansion effectively lifted the occupational floor for rural cohorts, enabling exit from low-skill manual labor into mid-tier employment. However, it did not dismantle the structural segmentation that reserves professional and managerial positions disproportionately for urban incumbents. Complementary demand-side reforms, such as hukou liberalization and anti-discrimination measures in formal-sector hiring, appear necessary to convert mid-tier upgrading into broad-based occupational convergence.

9. Conclusion

This paper estimates the effect of China's 1999 higher-education expansion on the urban–rural divide in occupational outcomes. Methodologically, it introduces an LLM-assisted roster-matching procedure that assigns respondents' occupation descriptions to an interpretable ten-tier socioeconomic scale, offering a replicable alternative to conventional occupation-status indices in settings where detailed task or wage data are unavailable.

Reduced-form difference-in-differences estimates show that the expansion enabled meaningful occupational upgrading for rural cohorts by raising status scores by 0.25 points (roughly 17% of the raw urban–rural gap) and reducing concentration in low-skill employment, but that convergence operates almost entirely within the lower half of the occupational hierarchy. A shift-share instrumental variable strategy confirms a causal education channel: each additional year of schooling raises occupational status by 0.21 points for rural compliers. Access to professional and managerial positions remains unchanged, and the equalizing power of a college degree for rural graduates has eroded as tertiary enrollment expanded toward mass participation.

These findings help explain a broader puzzle. China's college expansion is widely credited with raising human capital and narrowing urban–rural disparities, yet two decades after the reform the occupational divide between rural- and urban-origin workers remains conspicuous. Our results suggest that this persistence is not due to a failure of education to raise occupational status but rather to an institutional ceiling that confines education-driven gains to mid-tier employment. The same logic may apply to other developing countries where rapid tertiary expansion coexists with rigid labor-market segmentation: supply-side education policy can lift the occupational floor but is unlikely, by itself, to dismantle structural dualism.

The policy implication is specific: while expanding college access facilitates the transition of rural cohorts from low-skill manual labor into skilled trades, converting this upward mobility into professional and managerial careers requires substantive demand-side reforms. To bridge this gap, potential policy directions include relaxing hukou-based hiring preferences, institutionalizing anti-discrimination measures, and formalizing the recognition of credentials from non-elite universities that disproportionately serve rural students.

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, the authors used generative AI tools for two limited purposes: (i) to improve language clarity and readability in parts of the manuscript, and (ii) as part of the occupational-text scoring procedure described in the Methods and Appendix. All outputs were reviewed, verified, and, where necessary, edited by the authors. The authors take full responsibility for the content of the publication.

Funding

This research received no external funding.

Author Contributions

R.Z., Y.T., and Z.W. jointly designed the study and contributed to all major aspects of the research, including conceptualization, empirical analysis, and writing. All authors proofread the manuscript, revised it critically for important intellectual content, and approved the final version. R.Z., Y.T., and Z.W. contributed equally to this work and share first authorship.

References

Adão, R., Kolesár, M., & Morales, E. (2019, 11). Shift-share designs: Theory and inference. *The Quarterly Journal of Economics*, *134*, 1949-2010. Retrieved from https://dx.doi.org/10.1093/qje/qjz025 doi: 10.1093/qje/qjz025

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*(1), 727–753.

Athey, S., Brunborg, H., Du, T., Kanodia, A., & Vafa, K. (2024). Labor-llm: Language-based occupational representations with large language models. *arXiv preprint arXiv:2406.17972*.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, *118*(4), 1279–1333.

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. Chicago: University of Chicago Press.

Borusyak, K., Hull, P., & Jaravel, X. (2022, 1). Quasi-experimental shift-share research designs. *The Review of Economic Studies*, *89*, 181-213. Retrieved from https://dx.doi.org/10.1093/restud/rdab030 doi: 10.1093/restud/rdab030

Borusyak, K., Hull, P., & Jaravel, X. (2025). A practical guide to shift-share instruments. *Journal of Economic Perspectives*, *39*(1), 181–204.

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

Chan, K. W., & Buckingham, W. (2008). Is china abolishing the hukou system? *The China Quarterly*, *195*, 582–606.

De Chaisemartin, C., & d' Haultfoeuille, X. (2018). Fuzzy differences-in-differences. *The Review of Economic Studies*, *85*(2), 999–1028.

Duncan, O. D. (1961). A socioeconomic index for all occupations. In J. Reiss Albert J. (Ed.), *Occupations and social status* (pp. 109–138). New York: Free Press.

Finlay, K., & Magnusson, L. M. (2009). Implementing weak-instrument robust tests for a general class of instrumental-variables models. *The Stata Journal*, *9*(3), 398–421.

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social science research*, *21*(1), 1–56.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, *120*(30), e2305016120.

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

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020, 8). Bartik instruments: What, when, why, and how. *American Economic Review*, *110*, 2586-2624. doi: 10.1257/AER.20181047

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), 267–297.

Hudson, S., Hull, P., & Liebersohn, J. (2017). Interpreting instrumented difference-in-differences. *Metrics Note, Sept.*

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, *133*(1), 97–126.

Li, C. (2005). prestige stratification in the contemporary china: Occupational prestige measures and socio-economic index. *Sociological Study*, *2*, 74–102.

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.

Ministry of Education of the People's Republic of China. (2000). *China education statistics yearbook 1999.* People's Education Press. (Compiled by the Ministry of Education; used to document tertiary admissions and enrollment around the 1999 expansion)

Miyaji, S. (2024). Instrumented difference-in-differences with heterogeneous treatment effects. *arXiv preprint arXiv:2405.12083*.

National Bureau of Statistics of China. (2000). *Statistical communiqué of the people's republic of china on the 1999 national economic and social development.* Retrieved 2026-02-05, from https://www.stats.gov.cn/english/NewsEvents/200203/t20020329_25980.html (Communiqué dated 2000-02-28 (official PDF); English webpage posted 2002-03-29)

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic perspectives*, *25*(1), 159–184.

Spence, M. (1973, 8). Job market signaling. *The Quarterly Journal of Economics*, *87*, 355-374. Retrieved from https://dx.doi.org/10.2307/1882010 doi: 10.2307/1882010

Staiger, D. O., & Stock, J. H. (1994). *Instrumental variables regression with weak instruments.* National Bureau of Economic Research Cambridge, Mass., USA.

Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear iv regression.* National Bureau of Economic Research Cambridge, Mass., USA.

Treiman, D. J. (1977). *Occupational prestige in comparative perspective.* New York: Academic Press. doi: 10.1016/C2013-0-11617-0

World Bank. (2025). *World development indicators: School enrollment, tertiary (% gross) (se.ter.enrr) — world.* Retrieved 2025-09-22, from https://data.worldbank.org/indicator/SE.TER.ENRR?locations=1W (World Development Indicators (WDI). Data source: UNESCO Institute for Statistics (UIS) via UIS API. License: CC BY 4.0.)

Appendix A. Online Appendix

Appendix A.1. LLM Scoring Prompt (Chinese and English Versions)

职业社会经济地位评分 Prompt

你是一个职业社会经济地位评分专家。请为每个职业描述打出 1-10 的整数分。

核心原则：

请根据每条职业描述的**实际工作性质、技能要求和社会经济地位**进行综合判断。下方锚定表提供的是各分数段的**参考标杆**，帮助你理解每个分数对应的社会经济层次，而不是关键词匹配规则。同一个词出现在不同语境中，应根据实际情况给出不同分数。

例如：

- “司机” 在锚定表中对应 4 分，但 “三轮车运输司机” 的实际工作性质是三轮车营运，属于低技能体力劳动，应为 **2** 分。
- “书记” 在锚定表中对应 10 分，但 “村支书记” 是村级干部，应为 **5** 分。

锚定量表：

分数	社会经济层次	参考职业
1	边缘生存	捡破烂, 拾荒
2	低技能体力劳动	保洁员, 务农, 搬运工, 洗碗工
3	基础服务业	服务员, 保安, 保姆, 收银员
4	技术工人	电工, 厨师, 司机, 焊工
5	普通白领	文员, 业务员, 出纳, 前台
6	初级专业人员	小学教师, 护士, 会计, 幼师
7	中级专业/基层管理	中学教师, 主治医师, 工程师, 律师
8	高级专业/中层管理	副教授, 副主任医师, 高级工程师, 注册会计师
9	资深专业/高层管理	教授, 主任医师, 总经理, 厂长
10	顶层领导/企业主	董事长, 总裁, 校长, 书记

行政级别专项规则：

中国党政机关、事业单位中的职务，按行政级别打分：

级别	正职	副职	示例
科级	8	7	科长、乡镇长、县局局长
处级	9	8	处长、县长、县委书记、市局局长
厅局级及以上	10	9	厅长、市长、市委书记、省级领导

当职务未标注行政层级时（如单独出现 “书记”、“局长”、“主席”），默认按该职务**最高常见级别**打分：

- “书记” → 默认县委书记以上 → **10**
- “人大主席” → 默认县级人大常委会主席以上 → **10**
- “局长” → 可能科级也可能厅级，默认 → **9**

语境判断提示：

“处”、“科”、“部” 等字在不同组织中含义不同：

- 在**党政机关、事业单位**中，对应真实行政级别，适用上述行政级别规则。
- 在**企业、银行等机构**中，仅为内部部门名称，不对应行政级别，应按管理层级的实际地位打分。

对比示例：

- "县教育局局长" → 党政机关，科级正职 → **8**
- "银行信贷处处长" → 银行内部部门负责人 → **7**

输出格式：
每行一条，格式为：职业，分数
不需要理由，不需要表头。无法判断的条目标记为 NA。

Occupation Socioeconomic Status Scoring Prompt

You are an expert in occupational socioeconomic status scoring. Please assign an integer score from 1 to 10 for each occupation description.

Core Principle:

Evaluate each occupation description based on **actual work content, skill requirements, and socioeconomic standing**. The anchor table below provides **reference benchmarks** for each score level, helping you understand the socioeconomic stratum corresponding to each score. It is not a keyword-matching rule. The same term appearing in different contexts should receive different scores based on the actual situation.

Examples:

- "Driver" corresponds to 4 in the anchor table, but "tricycle transport driver" involves low-skill manual labor and should be scored **2**.
- "Secretary" corresponds to 10 in the anchor table, but "village party secretary" is a village-level cadre and should be scored **5**.

Anchor Scale:

Score	Socioeconomic Stratum	Reference Occupations
1	Marginal survival	Scavenging
2	Low-skill manual labor	Cleaner, farmer, porter, dishwasher
3	Basic service	Waiter, security guard, nanny, cashier
4	Skilled worker	Electrician, cook, driver, welder
5	Ordinary white-collar	Clerk, salesperson, cashier, receptionist
6	Junior professional	Primary school teacher, nurse, accountant
7	Mid-level professional/manager	Secondary school teacher, physician, engineer, lawyer
8	Senior professional/manager	Associate professor, senior engineer, CPA
9	Executive/senior management	Professor, chief physician, GM, factory director
10	Top leadership/business owner	Chairman, president, principal, party secretary

---

Administrative Rank Rules:

For positions in Chinese government agencies and public institutions, score according to administrative rank:

Rank	Full	Deputy	Examples
Section level	8	7	Section chief, township head, county bureau director
Division level	9	8	Division chief, county head, county party secretary
Bureau level+	10	9	Bureau director, mayor, municipal party secretary

When administrative level is not specified (e.g., standalone "secretary," "director," "chairman"), default to the **highest common level**:

- "Secretary" → default county party secretary above → **10**
- "People's Congress Chairman" → default county level above → **10**
- "Bureau Director" → could be section or bureau level, default → **9**

Context Interpretation:

Terms like "division", "section", and "department" have different meanings in different organizations:

- In **government agencies and public institutions**, they correspond to actual administrative ranks; apply the administrative rank rules above.
- In **corporations, banks, and other enterprises**, they are merely internal department names without corresponding administrative ranks; score based on actual managerial position.

Comparative examples:

- "County Education Bureau Director" → government agency, section-level full position → **8**
- "Bank Credit Division Chief" → internal department head in a bank → **7**

Output Format:

One entry per line, format: `occupation,score`

No explanations needed, no header. Mark uninterpretable entries as NA.

Given that we are dealing with a Chinese occupational roster, in practice, we use the Chinese version prompt above. The English version is translated for the convenience of our readers. For processing efficiency, we divided the roster into 24 batches of approximately 1,000 entries and processed them using identical prompts and model parameters for all batches.

The anchor scale was constructed drawing on established frameworks in occupational stratification research while adapting to the Chinese context.

We primarily referenced the International Socio-Economic Index of Occupational Status (ISEI) developed by Ganzeboom et al. (1992). The ISEI provides a continuous measure derived from optimal scaling of occupational categories, ranging from low-skill manual labor at the bottom to professional and managerial positions at the top. Our 10-point scale mirrors this hierarchical structure, with anchor occupations selected to represent distinct socioeconomic strata. To ensure the anchor occupations better reflect Chinese social structure, we also drew on C. Li (2005), which provides empirically grounded prestige rankings based on Chinese survey data. This informed our selection of locally relevant anchor occupations and our treatment of positions within China's administrative hierarchy. Although both ISEI and Li's index employ continuous scales (0–100 respectively), we adopted a 10-point scale. This reflects a deliberate trade-off between granularity and anchor reliability.

A finer-grained scale would require precise distinctions among similar occupations—for instance, specifying whether a "cook" should score 43 or 47 while a "driver" scores 45 or 49. Such distinctions are inherently arbitrary. Any inconsistencies in anchor scores would propagate through the LLM's scoring, potentially introducing systematic bias.

With a 10-point scale, we only need to ensure that anchor occupations are correctly assigned to their appropriate stratum—that cooks and drivers both belong to the skilled worker category (score 4), without ranking one above the other. This makes the scoring framework more robust: the model can reliably place occupations into the correct stratum based on clear reference points, rather than attempting fine distinctions that even domain experts might disagree on.

Appendix A.2. Additional Variation from Text-Level Occupational Scoring

Table A.5: Within-Code Score Variation: Examples from the Five Largest Occupation Codes

Code	Industry	N	Score	Example Description (Translated)
50102	Agriculture	73,522	2	Corn planting; rice farming
			4	Driver; agricultural technician worker
			5	Salesperson; agricultural product sales
			6	Agricultural extension staff
			7	Farm director; Party branch secretary
40102	Retail Trade	45,396	1	Waste collection
			3	Vegetable seller; shop clerk
			5	Sales representative
			7	Sales manager; product manager
			10	CEO
62901	Construction	14,400	2	Manual laborer; odd-job worker
			4	Bricklayer; rebar worker
			6	Cost estimator; engineering technician
			9	General manager; company executive
40302	Food & Beverage	12,585	2	Dishwasher
			3	Waiter/waitress
			5	Restaurant owner
			7	Hotel lobby manager
40202	Transportation	12,382	2	Brick hauler; cargo hauler
			4	Driver; taxi driver
			6	Driving instructor
			9	General manager

Note: Each row shows individuals who share the same five-digit occupation code but received different scores based on their free-text occupation descriptions. N is the number of individuals born 1970–1988 assigned to each code. The five codes shown are the five largest by population. Occupation descriptions were originally in Chinese; English translations are provided here.

Table A.6: Variance Decomposition of Occupation Scores: Code-Level vs. Text-Level Scoring

Number of distinct occupation codes	437
Variance decomposition	
Total variance (σ^2_{total})	2.452
Between-code ($\sigma^2_{\text{between}}$)	1.640 (66.9%)
Within-code (σ^2_{within})	0.812 (33.1%)
R^2 (η^2)	0.669
Misclassification under code-mean imputation	
Individuals with $\lvert\Delta\rvert \geq 1$	64,727 (19.2%)
Individuals with $\lvert\Delta\rvert \geq 2$	13,887 (4.1%)

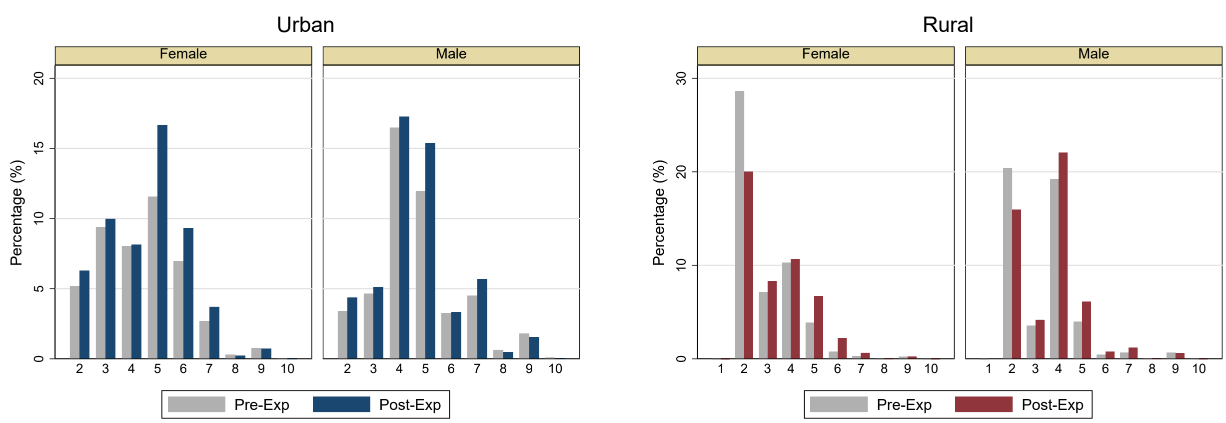
Note: This table reports a one-way ANOVA decomposition of occupation scores into between-code and within-code components. R^2 (η^2) is the share of total score variance explained by the five-digit occupation code alone. "Misclassification under code-mean imputation" counts individuals whose LLM-assigned score differs from their code's mean score by at least $\lvert\Delta\rvert$ points, representing the information loss from using a single code-level score instead of text-level LLM scoring.

Appendix A.3. Descriptive Statistics and Sample Composition

Table A.7: Descriptive Statistics by Cohort Group and Hukou Status

	Pre-Expansion (Birth Year 1970–1978)		Post-Expansion (Birth Year 1979–1988)	
	Urban	Rural	Urban	Rural
Age	41.192	41.363	31.249	31.189
	(2.589)	(2.565)	(2.885)	(2.913)
Male	0.478	0.535	0.461	0.541
	(0.500)	(0.499)	(0.498)	(0.498)
Minority	0.069	0.104	0.087	0.107
	(0.253)	(0.306)	(0.282)	(0.309)
Working	0.783	0.865	0.780	0.843
	(0.412)	(0.341)	(0.414)	(0.364)
Education level	4.267	2.881	4.750	3.404
	(1.608)	(0.775)	(1.718)	(1.137)
Occupation score	4.684	3.209	4.689	3.545
	(1.605)	(1.397)	(1.543)	(1.442)
N	72,694	121,513	85,799	124,640

Notes: Entries report sample means with standard deviations in parentheses. The sample is split into pre-expansion cohorts (birth years 1970–1978) and post-expansion cohorts (1979–1988), and by hukou status (Urban vs. Rural). *Working* is an indicator equal to one if the respondent is currently working (`is_working`=1). *Occupation score* is the LLM-based 1–10 occupational socioeconomic status measure; its sample size is smaller because some verbatim occupation descriptions are missing or cannot be matched. Variable-specific N's are reported in the bottom rows.



(a) Urban Cohorts

(b) Rural Cohorts

Figure A.5: Occupation Score Distribution by Gender: Pre- vs. Post-Expansion Cohorts
Notes: Panel (a) and (b) present the distributional shifts for urban and rural samples side-by-side. Inside each panel, the gender heterogeneity is displayed (Left bars: Female, Right bars: Male). Gray bars denote pre-expansion (1970–1978), and colored bars denote post-expansion (1979–1988).

Appendix A.4. Cohort Exposure Timing at the 1999 Expansion

Table A.8: 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 (Not clean)
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.

Appendix A.5. Provincial Expansion Intensity

Table A.9: 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.

Appendix A.6. Raw Cohort Trends in the LLM-Based Occupation Score

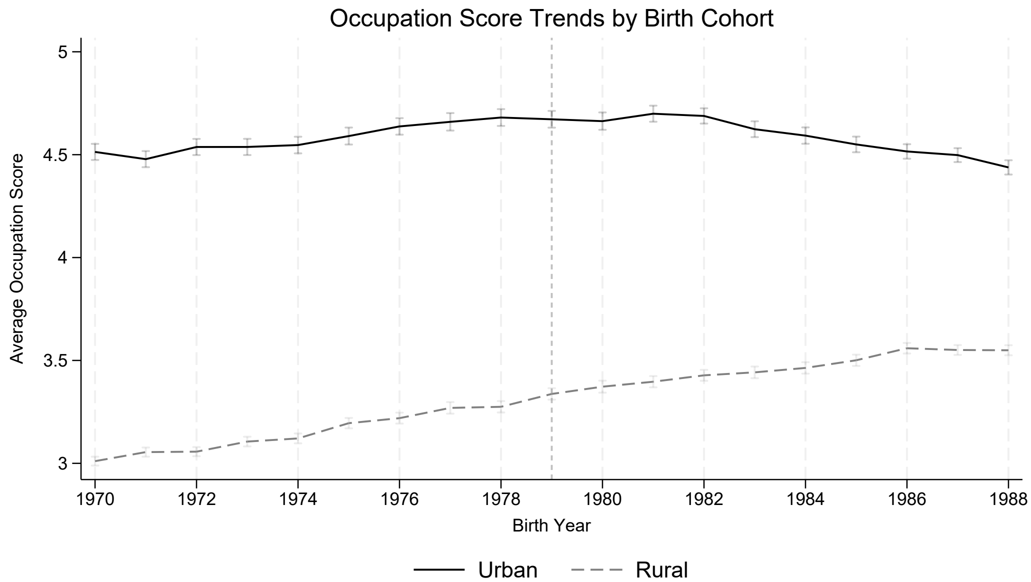


Figure A.6: Raw Trends in LLM-Based Occupation Score by Hukou Status (Urban vs. Rural)
Notes: This figure plots unadjusted (raw) cohort trends in the LLM-based occupation score separately for urban and rural respondents.

Appendix A.7. Distribution of Occupation Scores by Education and Cohort

Table A.10: Distribution of Occupation Scores by Education and Cohort (Row Percentages)

	Occupation Score									
Education Level	1	2	3	4	5	6	7	8	9	10
Panel A: Urban - Pre (1970-1978)										
Primary	0.22	38.51	12.97	33.54	12.25	0.72	0.53	0.03	0.47	0.77
Middle	0.11	14.80	16.25	39.55	22.40	1.94	1.83	0.14	1.23	1.75
High	0.03	4.28	12.41	33.28	32.92	8.58	4.32	0.44	2.03	1.72
College	0.00	0.59	3.26	10.63	34.58	28.53	15.97	2.38	3.12	0.94
Panel B: Urban - Post (1979-1988)										
Primary	0.00	49.33	10.38	29.44	8.97	0.65	0.50	0.00	0.29	0.43
Middle	0.07	20.58	14.62	39.70	20.55	1.78	1.05	0.04	0.53	1.09
High	0.01	5.71	13.97	31.31	34.10	7.99	3.62	0.19	1.63	1.48
College	0.00	0.94	4.46	11.05	40.17	24.50	14.93	1.13	2.07	0.74
Panel C: Rural - Pre (1970-1978)										
Primary	0.23	63.41	5.88	24.77	4.82	0.21	0.26	0.01	0.17	0.23
Middle	0.17	44.63	7.89	34.98	9.73	0.73	0.68	0.03	0.53	0.62
High	0.02	26.48	8.43	34.58	20.02	4.40	3.10	0.16	1.38	1.41
College	0.00	6.75	4.35	15.26	29.95	23.27	13.62	1.70	3.47	1.64
Panel D: Rural - Post (1979-1988)										
Primary	0.18	62.96	5.50	25.53	4.95	0.29	0.23	0.00	0.14	0.22
Middle	0.13	38.49	9.07	38.68	11.22	0.95	0.59	0.02	0.37	0.48
High	0.04	20.87	10.76	36.69	22.87	4.54	2.42	0.08	0.88	0.86
College	0.01	6.28	6.15	17.15	35.29	20.77	11.25	0.59	1.77	0.74

Note: Values represent row percentages (%). Each row sums to 100%.
Source: Calculated from survey data. The columns 1–10 represent the occupation score.

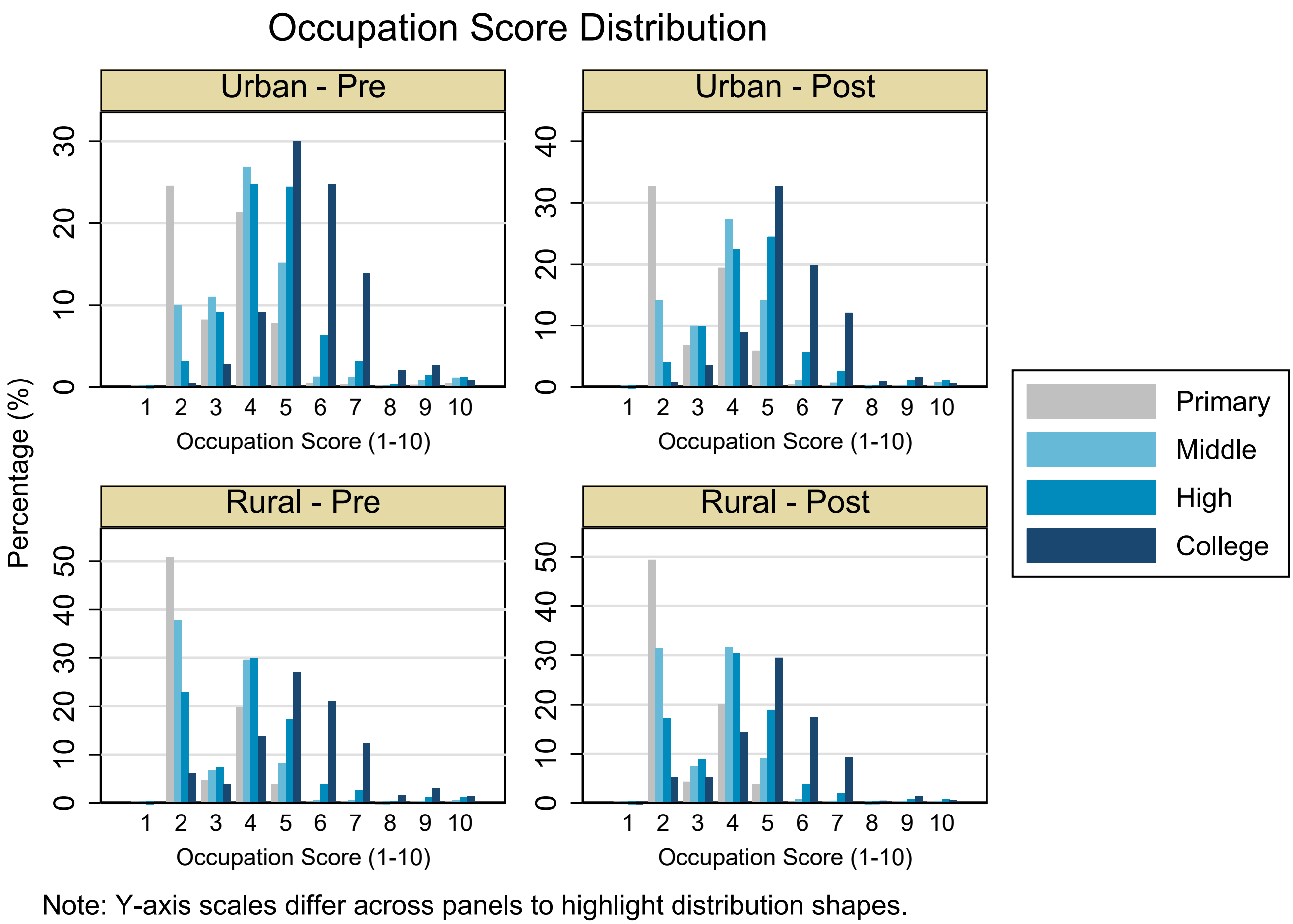


Figure A.7: Occupation Score Distribution by Education and Cohort

Note: Bars represent the percentage of individuals within each education level achieving a specific occupation score. Panels divide the sample by urban/rural status and pre/post-reform cohorts. Source: Calculated from survey data.

Appendix A.8. Additional Results: Occupational-Tier Effects

Table A.11: Effects of HE Expansion on Occupational-Tier Probabilities, by Gender

	Dependent variable: 1{Occupation score = k}								
	(1) Sc=2	(2) Sc=3	(3) Sc=4	(4) Sc=5	(5) Sc=6	(6) Sc=7	(7) Sc=8	(8) Sc=9	(9) Sc=10
Panel A: Full Sample (All Individuals)									
Female: Treatment (Post × Rural)	-0.113***	0.029***	0.024***	0.004	0.007***	0.000	0.002***	0.003***	0.000
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.000)	(0.001)	(0.000)
Observations	197,909	197,909	197,909	197,909	197,909	197,909	197,909	197,909	197,909
Male: Treatment (Post × Rural)	-0.088***	0.009***	0.056***	0.000	0.009***	-0.001	0.004***	0.006***	0.001***
	(0.003)	(0.003)	(0.005)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)	(0.000)
Observations	206,603	206,603	206,603	206,603	206,603	206,603	206,603	206,603	206,603
Panel B: Working-Only Sample (Employed Individuals)									
Female: Treatment (Post × Rural)	-0.121***	0.051***	0.047***	0.005	0.008**	0.000	0.003***	0.006***	0.000*
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)	(0.000)
Observations	140,772	140,772	140,772	140,772	140,772	140,772	140,772	140,772	140,772
Male: Treatment (Post × Rural)	-0.095***	0.011***	0.064***	-0.002	0.011***	-0.002	0.004***	0.008***	0.001***
	(0.004)	(0.003)	(0.005)	(0.004)	(0.002)	(0.003)	(0.001)	(0.002)	(0.000)
Observations	181,111	181,111	181,111	181,111	181,111	181,111	181,111	181,111	181,111
County & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each cell reports the coefficient on Post × Rural from a separate linear probability model (LPM). The dependent variable equals one if the individual' s LLM-based occupation score is k and zero otherwise; coefficients are percentage-point effects. Standard errors clustered at the county level are in parentheses. All specifications include county and year fixed effects and the full set of controls (as in the baseline specification). Panel A reports unconditional effects in the full sample. Individuals for whom an occupation score is not observed are not assigned to any tier $k \in 2, \ldots, 10$. Panel B restricts to individuals with a valid occupation score (employed). The occupation score is a 1–10 socioeconomic-status rank (1=lowest, 10=highest; see text for anchors). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix A.9. IV Estimates with the ISEI Outcome

Table A.12: The Causal Effect of Education on Occupational Status (ISEI Score)

	(1) OLS	(2) OLS	(3) First Stage	(4) Reduced Form	(5) 2SLS	(6) 2SLS
Panel A: Rural subsample						
Years of education	0.853***	0.853***			2.573**	2.573***
	(0.087)	(0.029)			(1.092)	(0.950)
Bartik IV			2.186***	5.625***		
			(0.707)	(1.989)		
Observations	103,856	103,856	103,856	103,856	103,856	103,856
R^2	0.137	0.137	0.265	0.115		
KP Wald F					9.55	29.17
Panel B: Urban subsample						
Years of education	2.490***	2.490***			2.193	2.193
	(0.048)	(0.030)			(2.214)	(2.585)
Bartik IV			0.945	2.072		
			(0.590)	(2.510)		
Observations	58,764	58,764	58,764	58,764	58,764	58,764
R^2	0.282	0.282	0.285	0.113		
KP Wald F					2.56	3.48
Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster (province)	Yes		Yes	Yes	Yes	
Cluster (county)		Yes				Yes

Notes: This table reports estimates of the effect of years of schooling on occupational status measured by the International Socio-Economic Index (ISEI). Panel A restricts the sample to individuals with rural *hukou*; Panel B to urban *hukou*. Columns (1)–(2) report OLS estimates; column (3) reports the first-stage effect of the Bartik instrument on years of education; column (4) reports the reduced-form effect on ISEI; columns (5)–(6) report 2SLS estimates. The Bartik instrument interacts province-level higher-education expansion intensity with county-level pre-reform high school enrollment share (1960–65 cohorts) and a post-reform cohort indicator (birth year $\geq$ 1979). All specifications absorb county and birth-year fixed effects and control for gender and minority status. Odd-numbered columns cluster standard errors at the province level (31 clusters); even-numbered columns cluster at the county level. KP Wald F is the Kleibergen–Paap rk Wald F statistic for weak identification. The sample includes cohorts born 1974–1986 from the 2015 1% Population Survey. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix A.10. Complementary outcome: Rural–urban convergence in low-skill employment

A key implication of occupational upgrading is that rural cohorts exposed to the higher-education expansion should become less likely to work in low-skill jobs, relative to comparable urban cohorts. Figure A.8 reports event-study estimates of the rural–urban gap in low-skill employment by birth cohort. Coefficients are cohort-by-rural interactions (normalized to zero for the 1978 cohort), with pointwise 95% confidence intervals; the vertical line marks the 1979 cohort cutoff. The pre-1979 estimates are close to zero and do not display a systematic trend, offering visual support for the absence of differential pre-trends. In contrast, the post-1979 cohorts exhibit a clear and persistent downward shift, indicating that rural cohorts become progressively less likely to be employed in low-skill occupations relative to urban cohorts.

Table A.13 summarizes these cohort dynamics using a baseline difference-in-differences specification. The coefficient on Post × Rural is negative and precisely estimated across all samples: −0.031 (s.e. 0.003) in the full sample, −0.040 (s.e. 0.004) for men, and −0.023 (s.e. 0.005) for women. Interpreted in levels, the estimates imply that the expansion reduced rural individuals' probability of working in a low-skill occupation by 3.1 percentage points relative to the corresponding change for urban individuals, with a larger reduction for men. In magnitude, this is economically meaningful given the high baseline prevalence of low-skill employment in our data (the fitted intercept is about 0.86): the implied decline corresponds to roughly 3.6% of the baseline level in the full sample (about 4.7% for men and 2.8% for women).

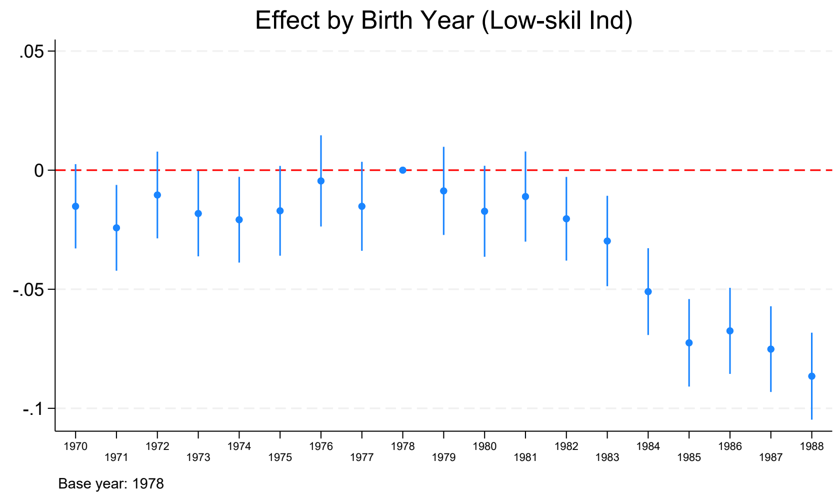


Figure A.8: Event-study estimates of the rural–urban gap in low-skill employment by birth cohort

Notes: The figure plots coefficients on cohort-by-rural interactions from the event-study specification, with pointwise 95% confidence intervals. The baseline cohort is 1978 (normalized to zero). Regressions include county fixed effects and county-by-rural fixed effects, controlling for gender and ethnicity; standard errors are clustered at the county level.

Table A.13: Baseline DID Estimates of Rural–Urban Convergence in Low-Skill Employment

	(1) Full sample	(2) Male	(3) Female
Post × Rural	-0.031***	-0.040***	-0.023***
	(0.003)	(0.004)	(0.005)
Constant	0.858***	0.853***	0.834***
	(0.005)	(0.005)	(0.005)
Observations	323,462	181,932	141,339
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,811	2,804	2,789

Notes: The dependent variable is *low_skill*, an indicator for low-skill employment. The coefficient on Post × Rural captures the differential change in low-skill employment for rural individuals relative to urban individuals in post-expansion cohorts. All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. The main effects of *post* and *rural* are absorbed by the fixed effects and therefore omitted. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix A.11. Heterogeneous analysis: agriculture dependence

Table A.14: Heterogeneity by Agricultural Dependence: Terciles of Pre-Period Agricultural Output Share

	(1) Full sample	(2) Male	(3) Female
Post × Rural	0.176***	0.158***	0.213***
	(0.021)	(0.026)	(0.030)
Post × Rural × Middle	0.175***	0.173***	0.173***
	(0.032)	(0.040)	(0.045)
Post × Rural × High	0.120***	0.169***	0.064
	(0.032)	(0.041)	(0.044)
Observations	274,084	154,457	119,491
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	Yes	Yes	Yes

Notes: Dependent variable is the LLM-generated occupation socioeconomic status score. The sample is split into terciles (Low/Middle/High) based on the *city-level* agricultural output share measured in the pre-period. The omitted group is the Low-tercile. All specifications include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. The main effects of *Post*, *Rural*, and tercile indicators are absorbed by the fixed effects and therefore omitted. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix A.12. Reduced-Form Robustness Checks

This section evaluates a set of robustness checks that speak to key threats to identification. We examine whether our baseline estimates are (i) sensitive to sample availability and measurement choices, (ii) driven by migration or hukou reclassification, (iii) driven by a subset of county-level administrative units, (iv) sensitive to cohort-window definitions and placebo tests, and (v) sensitive to inference choices such as the clustering level. Across these exercises, the estimated post-cohort rural–urban differential remains stable in sign and magnitude.

Appendix A.13. Robustness: Common-sample robustness check

Because ISEI is missing for a subset of observations, ISEI-based regressions are estimated on a smaller sample. To rule out compositional differences induced by sample availability, we re-estimate the LLM occupation-score specification on the overlapping sample with non-missing ISEI, holding fixed effects, controls, and the clustering scheme unchanged. Figure A.9 shows that the event-study profile and point estimates are essentially unchanged, suggesting that our conclusions are not driven by differential sample selection.

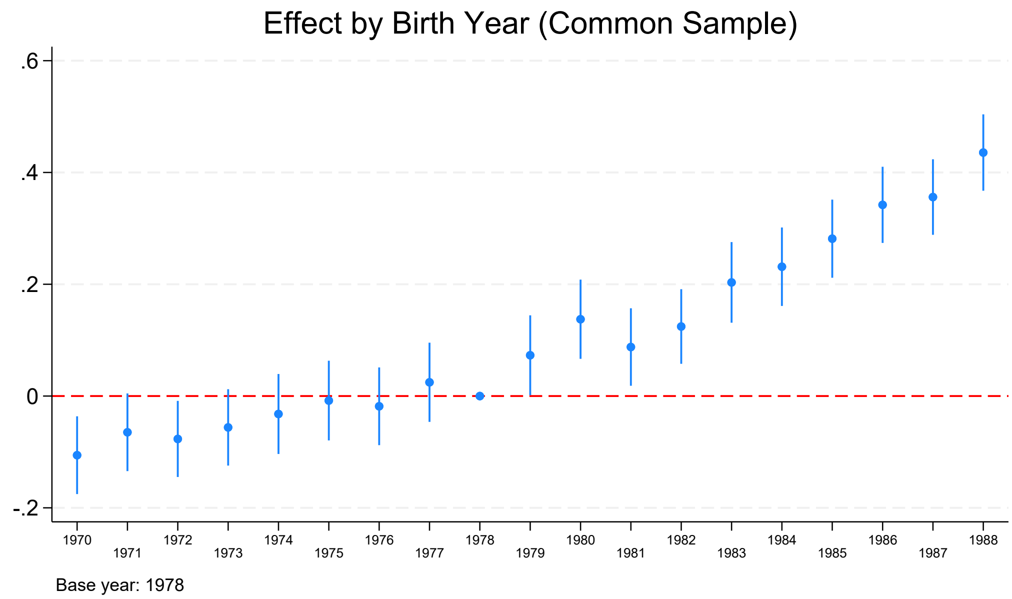


Figure A.9: Common-sample robustness: LLM occupation score estimated on the ISEI non-missing sample

Notes. Because ISEI is missing for a subset of observations, the ISEI regressions are estimated on a smaller sample. To ensure comparability, we re-estimate the LLM occupation-score specification on the common (overlapping) sample where ISEI is non-missing, keeping the same controls, fixed effects, and clustering as in the baseline.

Appendix A.14. Robustness: Hukou Reclassification and Migration

A concern is that hukou reclassification and migration could change the composition of rural and urban samples across cohorts, potentially confounding cohort-based comparisons. To address this, we restrict the sample to non-movers and reconstruct the migration-related outcome within this restricted sample. As shown in Figure A.10, the main patterns remain, indicating that our baseline results are not driven by differential migration or hukou-related compositional changes.

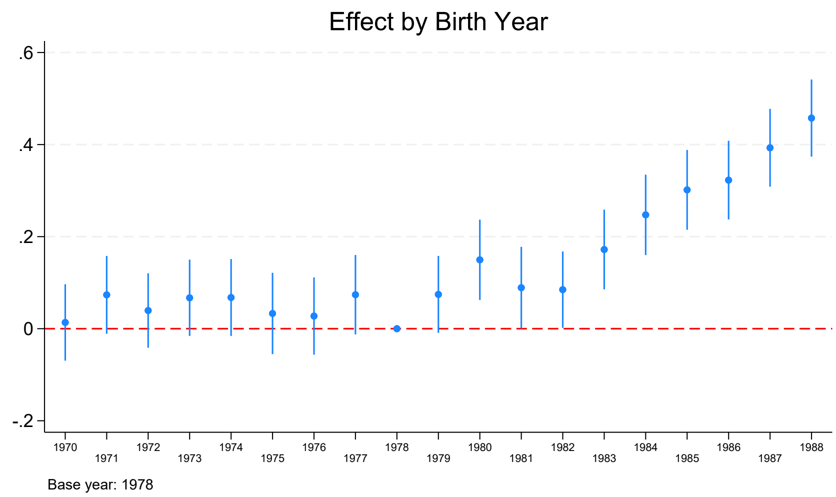


Figure A.10: Migration outcome among non-movers (restricted sample).
Notes: This figure restricts the sample to individuals who did not migrate (non-movers). The migration-related outcome is constructed within this restricted sample. See the text for variable definitions and sample construction details.

Appendix A.15. Robustness: Excluding Urban Districts

Municipal districts (*shixiaqu*) and direct-controlled municipalities may differ systematically from other county-level units in labor-market structure and policy exposure. We therefore drop municipal districts within prefecture-level cities and all direct-controlled municipalities. Table A.15 shows that the estimated relationship between education and the LLM occupation score is virtually unchanged, suggesting that our results are not driven by this subset of county-level administrative units.

Table A.15: Robustness Check: Education and LLM Occupation Score (Excluding Municipal Districts and Municipalities)

	(1) Full sample	(2) Male	(3) Female
LLM Occupation Score	0.444***	0.410***	0.480***
	(0.003)	(0.004)	(0.004)
Observations	302,016	169,956	131,868
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,738	2,731	2,716

Notes: Dependent variable is the LLM-generated occupation socioeconomic status score. This robustness check drops (i) municipal districts (*shixiaqu*) within prefecture-level cities and (ii) all direct-controlled municipalities. Importantly, municipal districts and counties are both county-level administrative units in China. Hence, this restriction removes a subset of county-level units (and municipalities) rather than excluding "urban areas" per se. All specifications include county fixed effects, birth-year fixed effects, and county×rural fixed effects; robust standard errors clustered at the county level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix A.16. Robustness: Placebo Cohorts (1970–1978)

To probe the identifying assumption, we implement placebo event studies using pre-reform cohorts (1970–1978) that should not be affected by the 1999 expansion. Figure A.11 shows no systematic placebo effects relative to the omitted cohort, lending support to the absence of differential pre-trends driving our baseline estimates.

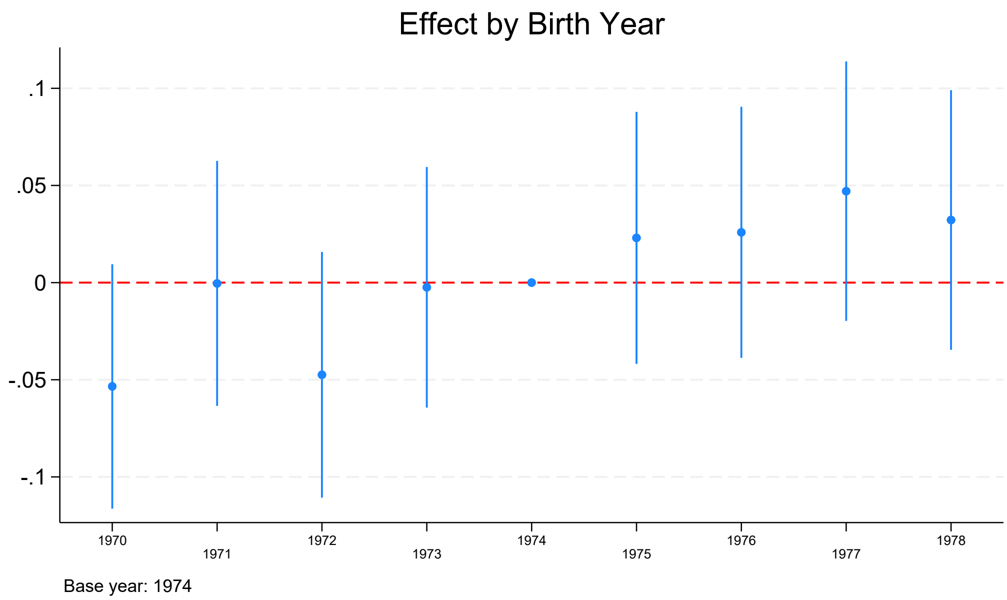


Figure A.11: Placebo Event Study: Pre-reform Cohorts (1970–1978)

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.

Appendix A.17. Robustness: Alternative Cohort Windows

As an additional check, we exclude cohorts born around the policy boundary (1978–1980) to reduce sensitivity to gradual roll-out or cohort misclassification. Table A.16 confirms that the post-cohort rural–urban differential remains robust.

Table A.16: Robustness: Dropping Border Cohorts (1978–1980)

	(1) Full sample	(2) Male	(3) Female
Post × Rural	0.286***	0.293***	0.289***
	(0.014)	(0.017)	(0.018)
Male	-0.287***		
	(0.006)		
Han ethnicity	0.157***	0.151***	0.160***
	(0.017)	(0.021)	(0.023)
Constant	3.943***	3.627***	3.406***
	(0.019)	(0.019)	(0.021)
Observations	277,451	156,066	121,158
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,806	2,799	2,783

Notes: Dependent variable is the LLM-generated occupation socioeconomic status score. The sample excludes the border cohorts born in 1978–1980. All specifications include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. The main effects of *Post* and *Rural* are absorbed by the fixed effects and therefore omitted. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix A.18. Robustness: Alternative Clustering Choices

Because our expansion-intensity measure is defined at the provincial level, shocks may be correlated within provinces, making county-clustered inference potentially too optimistic. We therefore re-estimate the 2SLS specifications with province-clustered standard errors. Table A.17 shows that the first stage remains strong and the second-stage estimates are similar in sign and magnitude, confirming that our conclusions are robust to more conservative inference.

Table A.17: The Effect of Education on Occupational Status (2SLS; Province-Clustered Standard Errors)

	(1) LLM Occupation Score	(2) ISEI Score
Years of Schooling	0.185**	3.071***
	(0.068)	(0.705)
Years of Schooling × Rural	0.161***	-0.249
	(0.022)	(0.210)
Province FE	Yes	Yes
Birth-year FE	Yes	Yes
County × Rural FE	Yes	Yes
Observations	322,076	255,558
Kleibergen–Paap rk Wald F	15.248	14.178
First-stage F (excluded IVs)	24.520	23.900

Notes: This table reports 2SLS estimates of the effect of years of schooling on occupational status. Standard errors (in parentheses) are robust and clustered at the province level. The endogenous regressors are years of schooling and its interaction with rural status. The excluded instruments are the Bartik shift–share instrument (baseline exposure × post) and its interaction with rural status. All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix A.19. Potential Confounders and Concurrent Shocks

Table A.18: Robustness Check: Accounting for the 1998 Asian Financial Crisis (High vs. Low Exposure Regions)

	(1) Full sample	(2) Male	(3) Female
Post × Rural	0.291***	0.245***	0.346***
	(0.030)	(0.040)	(0.042)
Post × High-exposure	0.012	-0.021	0.044
	(0.029)	(0.039)	(0.040)
Post × Rural × High-exposure	-0.038	0.028	-0.107**
	(0.033)	(0.043)	(0.046)
Observations	302,016	169,956	131,868
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,738	2,731	2,716

Notes: Dependent variable is the LLM-generated occupation socioeconomic status score. This table implements a triple-differences specification to address a potential confounder from the 1998 Asian Financial Crisis. Regions are classified into high- vs. low-exposure groups based on the pre-period financial-industry share, and we allow the post-reform rural–urban gap to differ by crisis exposure via the $Post \times Rural \times High$ interaction. All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. Main effects of $Post$, $Rural$, and the exposure indicator are absorbed by the fixed effects and therefore omitted. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.19: DDD Controlling for Exposure to the 1998 Flood (Flood-affected Counties = 1)

	(1) Full sample	(2) Male	(3) Female
Post × Rural	0.251***	0.255***	0.254***
	(0.012)	(0.016)	(0.017)
Post × Flood-affected county	-0.028	-0.068	-0.028
	(0.066)	(0.104)	(0.090)
Post × Rural × Flood-affected county	0.085	0.110	0.118
	(0.086)	(0.112)	(0.114)
Observations	322,076	181,111	140,772
R-squared	0.304	0.289	0.349
Within R-squared	0.014	0.003	0.003
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,811	2,804	2,789

Notes: Dependent variable is the LLM-based occupation score. This difference-in-difference-in-differences (DDD) specification interacts the post-reform cohort indicator, rural status, and an indicator for counties affected by the 1998 flood (flood-affected county = 1). All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. Main effects of the post indicator, rural status, and the flood indicator (as well as the rural×flood interaction) are absorbed by the fixed effects and therefore omitted. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.20: DDD Controlling for the Nine-Year Compulsory-Education Catch-up Campaign (Treated Counties = 1)

	(1) Full sample	(2) Male	(3) Female
Post × Rural	0.244***	0.247***	0.251***
	(0.013)	(0.016)	(0.017)
Post × Campaign (Treated county)	-0.171***	-0.151**	-0.188***
	(0.045)	(0.061)	(0.059)
Post × Rural × Campaign (Treated county)	0.159***	0.163**	0.145**
	(0.048)	(0.066)	(0.064)
Observations	322,076	181,111	140,772
R-squared	0.304	0.289	0.349
Within R-squared	0.014	0.003	0.003
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,811	2,804	2,789

Notes: Dependent variable is the LLM-based occupation score. The specification is a difference-in-difference-in-differences (DDD) design that interacts the post-reform cohort indicator, rural status, and an indicator for counties exposed to the nine-year compulsory-education catch-up campaign (treated counties = 1). All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. Main effects of the post indicator, rural status, and the campaign indicator (as well as the rural×campaign interaction) are absorbed by the fixed effects and therefore omitted. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.21: DDD Controlling for SOE Restructuring Exposure (High vs. Low SOE Employment Share in the Pre-period)

	(1) Full sample	(2) Male	(3) Female
Post × Rural	0.291***	0.245***	0.346***
	(0.030)	(0.040)	(0.042)
Post × High SOE exposure	0.023	-0.007	0.051
	(0.029)	(0.039)	(0.039)
Post × Rural × High SOE exposure	-0.046	0.015	-0.108**
	(0.033)	(0.043)	(0.046)
Observations	322,076	181,111	140,772
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,811	2,804	2,789

Notes: Dependent variable is the LLM-based occupation score. This difference-in-difference-in-differences (DDD) specification accounts for differential exposure to SOE restructuring by interacting the post-reform cohort indicator, rural status, and an indicator for provinces with high pre-period SOE employment shares (high SOE exposure = 1). All models include county fixed effects, birth-year fixed effects, and county×rural fixed effects; standard errors are clustered at the county level. Main effects of the post indicator, rural status, and the SOE exposure indicator (as well as the rural×SOE interaction) are absorbed by the fixed effects and therefore omitted. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.22: Robustness: Controlling for WTO Exposure (High vs. Low) (DDD)

	(1) Full sample	(2) Male	(3) Female
PostCohort × Rural	0.257***	0.278***	0.244***
	(0.017)	(0.023)	(0.023)
PostCohort × HighWTO	0.018	0.024	0.012
	(0.022)	(0.028)	(0.029)
PostCohort × Rural × HighWTO	-0.009	-0.041	0.026
	(0.025)	(0.031)	(0.033)
Observations	322,076	181,111	140,772
County FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
County × Rural FE	Yes	Yes	Yes
Clusters (county)	2,811	2,804	2,789

Notes: This table examines whether the estimated post-cohort rural–urban gap in the LLM-based occupation score is confounded by WTO-related trade exposure. *HighWTO* is an indicator for provinces above the sample-median WTO exposure (constructed from pre-period trade exposure). All specifications include county fixed effects, birth-year fixed effects, and county×rural fixed effects; robust standard errors are clustered at the county level. Main effects of *PostCohort*, *Rural*, and *HighWTO* are absorbed by fixed effects and omitted. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix A.20. Instrument Strength and Monotonicity-Supporting Evidence

Appendix Table A.23 reports heterogeneous first-stage estimates from our instrumented difference-in-differences (DID-IV) design. We present these subgroup regressions to assess the plausibility of the monotonicity restriction—namely, that stronger exposure to the expansion shock does not reduce schooling for any subpopulation (i.e., ruling out systematic "defiers"). While monotonicity is not directly testable, a necessary supporting pattern is that the instrument predicts schooling in the same direction across pre-specified groups. The results are consistent with this requirement: the estimated first-stage coefficients are positive across all demographic subsamples (by gender, hukou status, and their interaction) and statistically significant in most cases, with cluster-robust first-stage F statistics generally above conventional thresholds. Although the first stage is weaker in splits based on expansion intensity and baseline high-school share (Panel B), the estimates remain positive, suggesting that the expansion-induced increase in educational attainment operates broadly rather than being driven by a narrow subset of the population. Overall, the evidence in Table A.23 supports the monotonicity assumption and reinforces the credibility of the DID-IV strategy.

Table A.23: Heterogeneous First-Stage Estimates (Consistency with Monotonicity)

Panel A: By gender and hukou status

	Full	Male	Female	Urban	Rural
DID-IV instrument	2.370***	1.783***	2.894***	2.180***	3.596***
	(0.453)	(0.459)	(0.550)	(0.605)	(0.830)
First-stage F	27.3	15.1	27.7	13.0	18.8
Observations	322,236	181,258	140,957	118,198	203,998

Panel B: By expansion intensity and baseline high-school share

	High expansion intensity	Low expansion intensity	High baseline HS share	Low baseline HS share
DID-IV instrument	1.644**	2.158***	1.308**	2.492
	(0.678)	(0.604)	(0.540)	(1.759)
First-stage F	5.9	12.7	5.9	2.0
Observations	155,255	166,981	160,951	161,285

Panel C: By gender × hukou status

	Male ×Urban	Male ×Rural	Female ×Urban	Female ×Rural
DID-IV instrument	1.733**	2.937***	2.531***	4.524***
	(0.766)	(0.809)	(0.718)	(1.017)
First-stage F	5.1	13.2	12.4	19.8
Observations	61,879	119,295	56,213	84,618

Notes: This appendix table reports heterogeneous first-stage estimates to assess the plausibility of the monotonicity assumption (i.e., the absence of defiers) in an instrumented difference-in-differences (DID-IV) design. Each column reports a separate first-stage regression of years of schooling on the DID-IV instrument, estimated in pre-specified subsamples.

The instrument is the interaction of (i) province-level higher-education expansion intensity (measured over 1998–2006), (ii) a county-level baseline high-school (or above) enrollment share constructed from the 1960–1965 birth cohorts, and (iii) an indicator for post-expansion birth cohorts (birth year $\geq$ 1979).

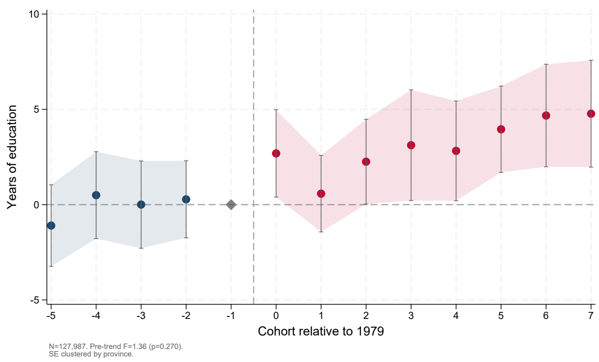
All specifications include county fixed effects and birth-year fixed effects. Controls include gender, minority status, and rural hukou status when applicable; when the sample is restricted by a covariate (e.g., male-only, urban-only), that covariate is omitted. Standard errors (in parentheses) are clustered at the province level. First-stage F is the cluster-robust F statistic for the null that the coefficient on the DID-IV instrument equals zero in the corresponding subsample. Monotonicity is not directly testable; however, the first-stage estimates are uniformly positive across the pre-specified subgroups reported here, which is consistent with the monotonicity restriction.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

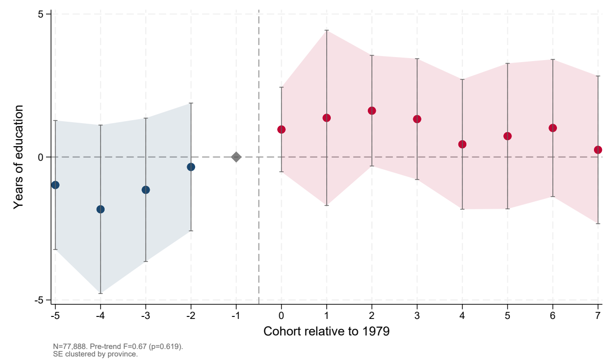
Appendix A.21. Pre-trends and No-Anticipation: First Stage and Reduced Form

To validate the parallel trends assumption required for our DID-IV design, we follow Hudson et al. (2017) and Miyaji (2024) in testing pre-trends separately for both the endogenous treatment variable (education) and the outcome variable (occupation scores). Figure A.12 presents event-study estimates for both the first stage and reduced form, separately by rural and urban subsamples. For our identification strategy to be valid, coefficients in the pre-reform period should be jointly indistinguishable from zero, indicating no differential trends between high- and low-exposure counties prior to the 1999 expansion. Joint F-tests confirm this requirement: we cannot reject that all pre-reform coefficients equal zero ($p > 0.10$ in all four panels), validating both components of the DID-IV parallel trends assumption.

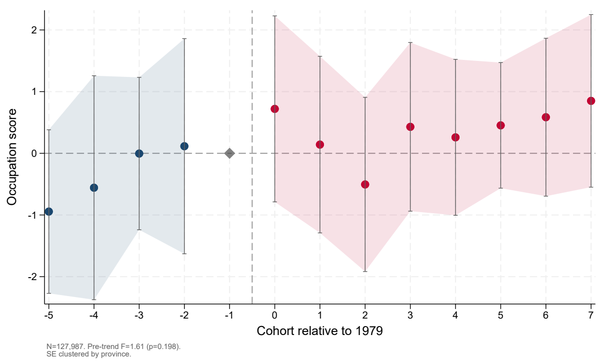
We restrict attention to adjacent birth cohorts (1974–1986) to keep comparisons local around the 1999 higher-education expansion: because college entry timing varies (school-entry age, repetition, and retakes), we treat the 1979 birth cohort as the first plausibly exposed cohort and use nearby cohorts as close pre-policy benchmarks. This narrow window also limits the scope for slow-moving county-specific changes in schooling—such as the gradual rollout of nine-year compulsory education—to accumulate into differential cohort trends, and it enables standard DID–IV event-study diagnostics for both the endogenous treatment (schooling) and the outcome (occupation scores). Predicted exposure is constructed using a predetermined county share: we measure county high-school completion from the 1960–1965 cohorts, who completed secondary schooling well before the expansion, and we average over multiple cohorts to reduce sampling noise. Finally, Appendix A.26 shows that our estimates and first-stage diagnostics are not sensitive to the exact baseline cohort window used to construct the county share (Appendix Figures A.14–A.16).



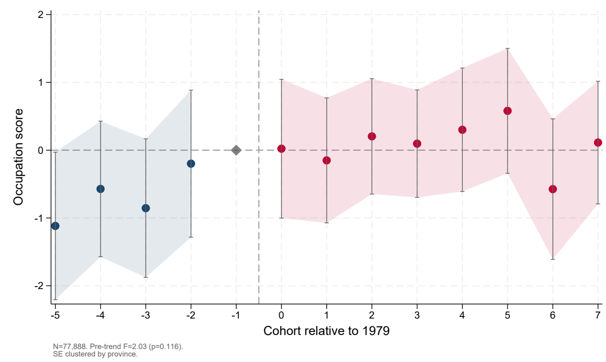
(a) First Stage: Rural



(b) First Stage: Urban



(c) Reduced Form: Rural



(d) Reduced Form: Urban

Figure A.12: Event-Study Estimates: First Stage and Reduced Form

Notes: Each panel plots event-study coefficients from regressions of the outcome (years of education in panels (a)–(b); occupation scores in panels (c)–(d)) on the interaction between predicted exposure intensity and birth cohort indicators. The instrument is constructed as $\text{Exposure}_c = \text{ExpProv}_p \times \text{Share}_c$, where province-level expansion intensity interacts with county-level high school completion rates from the 1960–1965 cohorts. The sample includes birth cohorts 1974–1986. Navy dots represent pre-reform cohorts (born 1974–1978), cranberry dots represent post-reform cohorts (born 1979–1986), and the gray diamond marks the reference cohort (born 1978, omitted category). The vertical dashed line at -0.5 separates pre- and post-reform periods. Shaded areas show 95% confidence intervals with standard errors clustered at the province level. All regressions include county and birth-year fixed effects and control for gender and minority status.

Appendix A.22. Testing for Anticipation Effects

Our difference-in-differences IV strategy requires that treated cohorts did not adjust their educational choices in anticipation of the 1999 expansion. This assumption is particularly important in our setting because the instrument interacts province-level expansion intensity with pre-reform county-level high school completion shares. If households or local governments anticipated the expansion and adjusted schooling decisions before 1999, the exclusion restriction could be violated through two channels: first, the first-stage relationship between predicted exposure and actual education could emerge prior to the reform; second, the county-level shares themselves could be endogenous to expected expansion intensity.

Table A.24 reports four complementary tests of the no-anticipation assumption. Panel A estimates placebo first-stage regressions using only pre-policy birth cohorts (1974–1978) with artificially early cutoffs at 1976 and 1977. If anticipation were driving the first stage, we would expect to see significant coefficients on predicted exposure even in these pre-reform samples. The placebo estimates are small and statistically insignificant in seven of eight specifications. The one marginal exception (rural, cutoff 1976, $p = 0.073$) is substantially smaller than the baseline first-stage coefficient and disappears when the cutoff moves to 1977.

For comparison, the actual first-stage coefficients using the correct 1979 cutoff and full sample are 3.245 (rural, $p < 0.001$) and 1.806 (urban, $p = 0.026$). This pattern suggests the first-stage relationship emerged after the reform rather than reflecting pre-existing trends.

Panel B scans alternative post-reform cutoffs from 1976 through 1983. In rural areas, the first-stage coefficient is stable across all cutoffs, ranging from 2.62 to 3.28 and remaining highly significant throughout. This stability is consistent with a sharp policy change in 1979 that affected all subsequent cohorts similarly. In urban areas, the coefficient is largest and most significant for cutoffs near the actual reform year (1976–1979) and declines monotonically for later cutoffs, eventually becoming insignificant by 1982. This gradient likely reflects the fact that urban college access expanded less dramatically and may have saturated more quickly, rather than indicating anticipation.

Panel C tests whether the county-level high school completion shares used to construct predicted exposure are themselves endogenous to province-level expansion intensity. We regress the county share (computed from 1960–1965 birth cohorts) on province expansion intensity at the county level, clustering standard errors by province. The coefficients are small and statistically insignificant in both rural ($\beta = 0.195$, $p = 0.415$) and urban ($\beta = 0.330$, $p = 0.382$) counties. This orthogonality is critical because the shares are predetermined by construction but could still be correlated with expansion intensity if provinces with higher planned expansion historically invested more in secondary schooling.

Panel D reports joint F-tests for pre-trends in the first-stage event study. We estimate event-study specifications interacting predicted exposure with indicators for each birth cohort relative to 1978 (the reference cohort) and jointly test whether the four pre-reform coefficients (cohorts 1974–1977) equal zero. The null hypothesis cannot be rejected in either sample (rural: $F(4, 30) = 1.36$, $p = 0.270$; urban: $F(4, 30) = 0.67$, $p = 0.619$), consistent with parallel trends in educational attainment prior to the expansion.

Table A.24: No-Anticipation Robustness Checks

	Rural	Urban
Panel A: Placebo First Stage (Pre-Policy Cohorts, 1974–1978)		
Predicted exposure ($c = 1976$)	0.949*	1.160
	(0.511)	(0.723)
Predicted exposure ($c = 1977$)	0.683	1.290
	(0.751)	(0.779)
Predicted exposure ($c = 1979$, baseline)	3.245***	1.806**
	(0.837)	(0.770)
Observations (placebo / baseline)	50,261 / 127,987	28,237 / 77,888
Panel B: First-Stage Coefficient by Alternative Post-Reform Cutoff		
$c = 1976$	2.620***	1.995**
	(0.760)	(0.755)
$c = 1977$	2.774***	2.071**
	(0.878)	(0.838)
$c = 1978$	2.918***	1.936**
	(0.846)	(0.819)
$c = 1979$ (baseline)	3.245***	1.806**
	(0.837)	(0.770)
$c = 1980$	2.866***	1.489*
	(0.808)	(0.790)
$c = 1981$	3.245***	1.154
	(0.903)	(0.755)
$c = 1982$	3.275***	0.753
	(0.911)	(0.739)
$c = 1983$	3.176***	0.387
	(0.894)	(0.717)
Panel C: Share Pre-Determination (County-Level Regression)		
Province expansion intensity	0.195	0.330
	(0.236)	(0.372)
Counties	1,716	1,096
Panel D: Pre-Trend Joint F-Test (First-Stage Event Study)		
$F(4, 30)$	1.362	0.668
p-value	[0.270]	[0.619]

Notes: This table reports tests of the no-anticipation assumption underlying our DID-IV design. Panel A estimates placebo first-stage regressions on pre-policy cohorts (1974–1978) using artificially early post-reform cutoffs ($c = 1976$ and $c = 1977$). The baseline row uses the actual cutoff ($c = 1979$) and full sample (1974–1986) for comparison. Panel B scans alternative cutoffs from 1976 to 1983 in the full sample. Panel C regresses county-level high school completion shares (computed from 1960–1965 cohorts) on province expansion intensity, testing whether the shares are predetermined. Panel D reports joint F-tests for pre-reform leads in first-stage event studies (testing whether coefficients on cohorts 1974–1977 jointly equal zero, relative to reference cohort 1978). All regressions include county and birth-year fixed effects and control for gender and minority status. Standard errors (in parentheses) are clustered at the province level (31 clusters). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix A.23. Testing the Exclusion Restriction

Table A.25 reports three tests of the exclusion restriction. Panel A adds province-by-birth-cohort fixed effects, absorbing all province-level time-varying confounders; identification relies solely on within-province variation in county-level high school completion shares. The rural 2SLS estimate (0.146) remains positive and broadly consistent with the baseline (0.214), though it loses precision as expected given the more demanding fixed-effect structure (KP $F = 8.59$). The decline in the point estimate may partly reflect the loss of identifying power from absorbing the province-level shift component; a wider Anderson-Rubin confidence interval that includes both the baseline estimate and zero would be consistent with this interpretation.

The urban estimate under these demanding fixed effects (0.346) is virtually unchanged from the baseline, though the very weak first stage (KP $F = 2.11$) renders it unreliable for inference. Panel B restricts to non-migrants, yielding estimates nearly identical to the baseline in both subsamples (rural: 0.215 vs. 0.214; urban: 0.344 vs. 0.342), with first-stage strength also preserved (rural KP $F = 15.13$; urban KP $F = 5.40$). The negligible change from excluding cross-county migrants rules out spatial sorting as an alternative channel. Panel C shows the instrument does not predict cross-county migration in either subsample (rural reduced-form coefficient: 0.006, $p > 0.10$; urban: 0.001, $p > 0.90$), confirming migration is not a pathway through which the instrument affects outcomes independently of education. Together, these tests support the identifying assumption that the instrument affects occupation scores only through education.

Table A.25: Exclusion Restriction Robustness Tests

	Rural		Urban	
Panel A: Province × Birth-Cohort Fixed Effects				
	First Stage	2SLS	First Stage	2SLS
Bartik IV / Edu years	2.414***	0.146	1.525	0.346***
	(0.823)	(0.154)	(1.049)	(0.116)
KP Wald F-stat	8.59		2.11	
Observations	127,987		77,888	
Panel B: Non-Migrant Subsample				
	First Stage	2SLS	First Stage	2SLS
Bartik IV / Edu years	3.250***	0.215**	1.816**	0.344***
	(0.835)	(0.104)	(0.782)	(0.064)
KP Wald F-stat	15.13		5.40	
Observations	127,941		77,733	
Panel C: Migration as Outcome				
	Reduced Form	2SLS	Reduced Form	2SLS
Bartik IV / Edu years	0.0055	0.0018	0.0006	0.0004
	(0.0036)	(0.0014)	(0.0067)	(0.0036)
Observations	154,234		103,673	

Notes: Panel A adds province-by-birth-cohort fixed effects to the baseline specification, absorbing all province-level time-varying confounders. Identification relies solely on within-province variation in county-level high school completion shares. Panel B restricts to individuals whose hukou registration matches their current county of residence. Panel C uses cross-county migration (indicator for whether current county differs from hukou registration county) as the outcome variable. The reduced form tests whether the instrument directly predicts migration; the 2SLS instruments education to test whether any migration effect operates through education. All specifications include county and birth-year fixed effects (Panel A uses province-by-cohort fixed effects instead of cohort fixed effects) and control for gender and minority status. Standard errors (in parentheses) are clustered at the province level (31 clusters). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

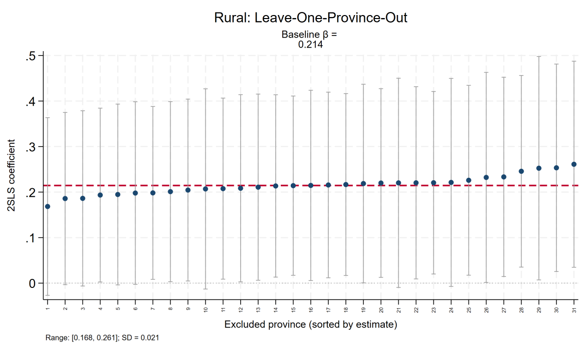
Appendix A.24. Robustness to Influential Provinces

Because the shift component of our Bartik IV varies across only 31 provinces, a single influential province could in principle drive the 2SLS estimate. We conduct a leave-one-province-out jackknife exercise, sequentially dropping each province and re-estimating the baseline specification. As shown in Figure A.13, the rural estimates range narrowly from 0.168 to 0.261 (SD =0.02) around the baseline of 0.214, and the urban estimates from 0.315 to 0.371 (SD =0.013) around 0.342. No single province drives the sign or magnitude of the effect.

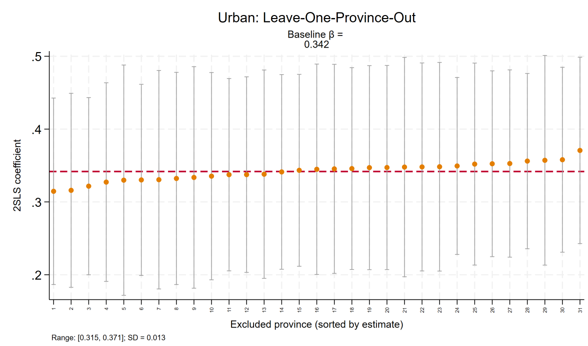
Appendix A.25. Leave-One-Province-Out Sensitivity

A natural concern with shift-share instruments is that the estimated effect could be disproportionately driven by a single influential province. Because the shift component of our Bartik IV—province-level expansion intensity—varies across only 31 provinces, even one province with an idiosyncratic shock or an unusually strong first-stage relationship could mechanically anchor the 2SLS estimate (Borusyak et al., 2025; Goldsmith-Pinkham et al., 2020). To assess this sensitivity, we conduct a leave-one-province-out (LOPO) jackknife exercise, re-estimating the baseline 2SLS specification while sequentially excluding each of the 31 provinces.

Figure A.13 reports the results. In the rural subsample (Panel a), the 31 LOPO estimates range from 0.168 to 0.261 with a standard deviation of only 0.021, tightly clustered around the full-sample estimate of 0.214. No single province's exclusion shifts the coefficient outside the baseline confidence interval, and all 31 estimates remain positive. The urban estimates (Panel b) are even more stable, ranging from 0.315 to 0.371 (SD = 0.013) around the baseline of 0.342. These patterns confirm that our results are not reliant on any particular province-level shock.



(a) Rural subsample



(b) Urban subsample

Figure A.13: Leave-One-Province-Out Sensitivity of the 2SLS Estimate

Notes: Each point represents the 2SLS estimate of the effect of years of education on occupation scores after excluding one province from the sample. Vertical bars denote 95% confidence intervals. The red dashed line marks the full-sample baseline estimate ($\hat{\beta} = 0.214$ for rural; $\hat{\beta} = 0.342$ for urban). Provinces are sorted by the magnitude of the leave-one-out estimate along the horizontal axis. All specifications use the baseline Bartik IV (province expansion intensity $\times$ county high school share $\times$ post-reform cohort indicator) with county and birth-year fixed effects, controlling for gender and minority status. Standard errors are clustered at the province level.

Appendix A.26. Sensitivity to the Baseline Share Window

This appendix examines whether our DID–IV results and diagnostics are sensitive to the choice of the baseline cohort window used to construct the county exposure share (high-school completion rate). We re-compute the share using 55 alternative baseline windows (start years 1960–1972; widths 2–6 cohorts), and report how the 2SLS estimate, instrument strength, and first-stage pre-trend diagnostics vary across these choices.

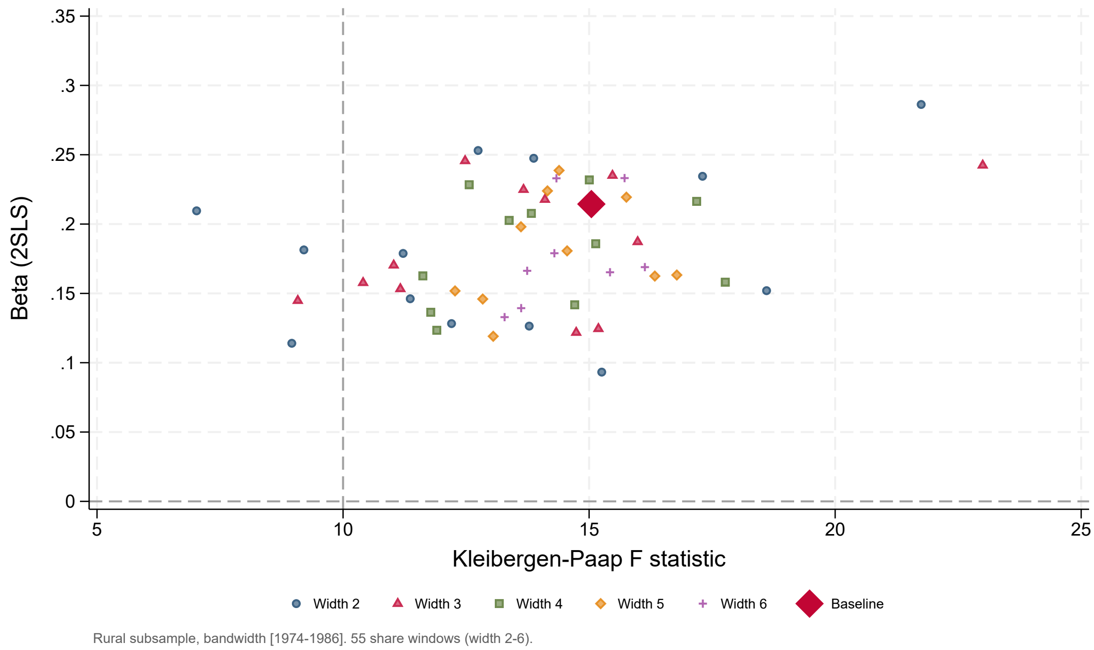


Figure A.14: Sensitivity to the Baseline Share Window: 2SLS Estimate vs. Instrument Strength (Rural)
Notes: Each point corresponds to an alternative baseline cohort window used to construct the county share (high-school completion rate), which is then interacted with province-level expansion intensity to form predicted exposure. The vertical axis reports the 2SLS estimate of the return to schooling, and the horizontal axis reports the Kleibergen–Paap weak identification statistic. Marker shapes/colors indicate the width of the baseline window (2–6 cohorts). The red diamond denotes our baseline choice (share constructed from the 1960–1965 cohorts). Rural subsample; main bandwidth [1974–1986].

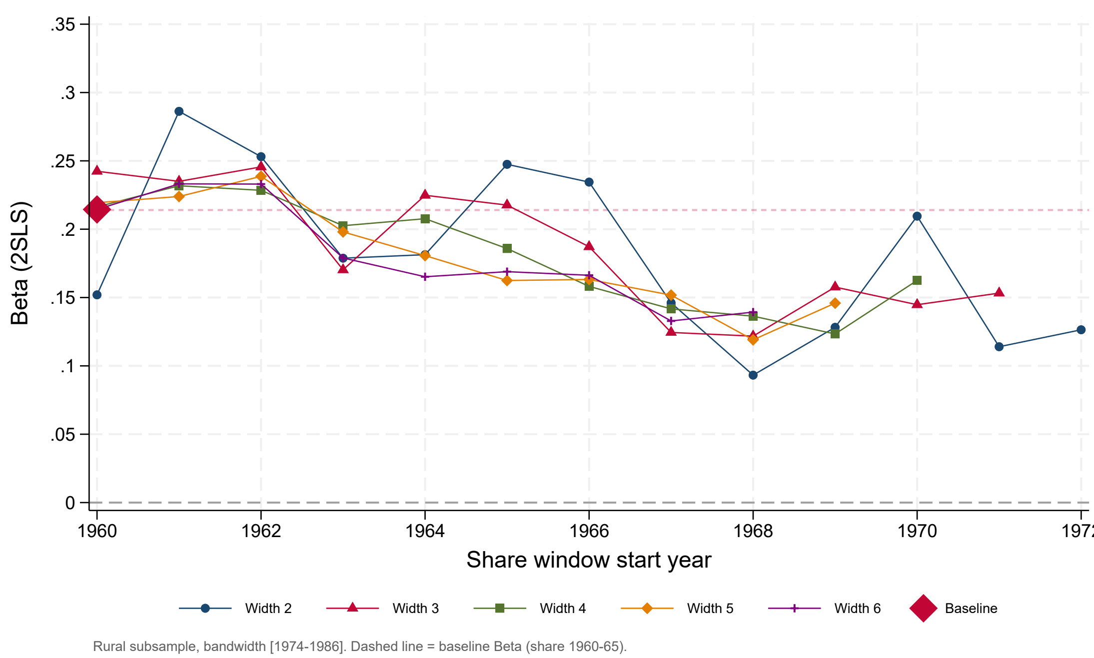


Figure A.15: Sensitivity to the Baseline Share Window: 2SLS Estimate by Baseline Start Cohort (Rural)
Notes: The figure plots the 2SLS estimate as a function of the *start year* of the baseline cohort window used to construct the county share. Separate series correspond to different window widths (2–6 cohorts). The red diamond marks the baseline specification using the 1960–1965 cohorts; the dashed horizontal line indicates the baseline 2SLS estimate. Rural subsample; main bandwidth [1974–1986].

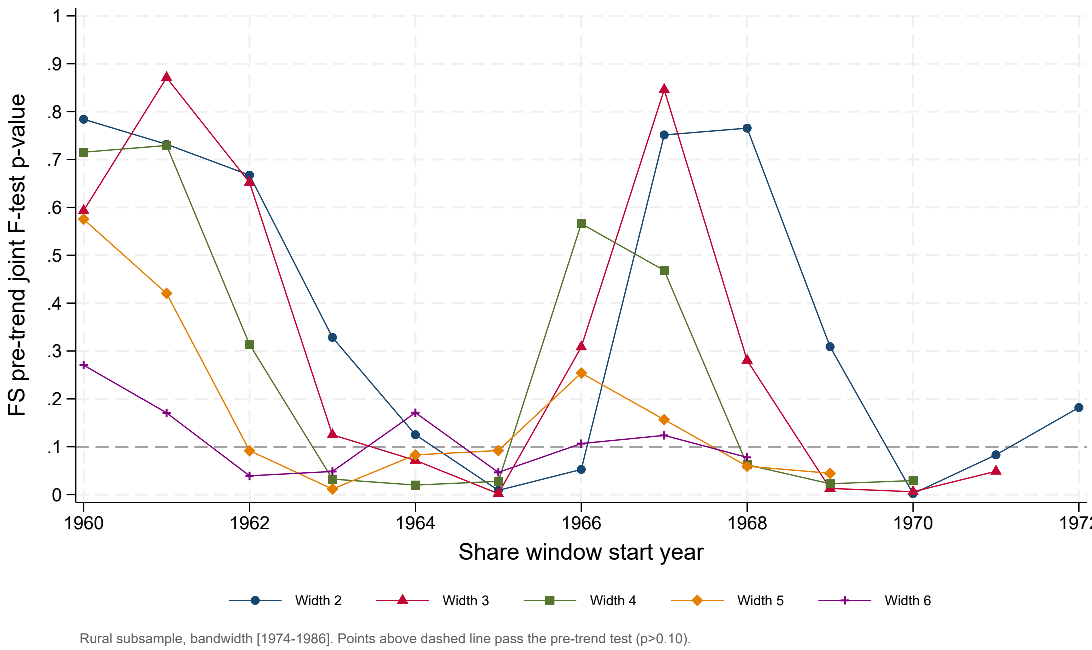


Figure A.16: Sensitivity to the Baseline Share Window: First-Stage Pre-trend Test by Baseline Start Cohort (Rural)
Notes: This figure reports the p-value from the *joint* pre-trend F-test in the first-stage event-study regression (testing whether all pre-policy lead coefficients are jointly zero), plotted against the start year of the baseline cohort window used to construct the county share. Separate series correspond to different baseline window widths (2–6 cohorts). Points above the dashed line ($p = 0.10$) pass the pre-trend diagnostic. Rural subsample; main bandwidth [1974–1986].