Industrial Places and Young Singlehood:

Cross-Sections and Cohort Histories from China's Third Front*

Yi Yao† Tai Lo Yeung‡

30 August 2026

Abstract

Places industrialized by administrative allocation carry a distinctive demographic profile. In China's 1982 census, the never-married contrast between ages 20–24 and 30–34 steepens by 4.23 percentage points per standard deviation of industrial composition inside Third Front provinces relative to elsewhere; predetermined siting controls reduce it to 2.24 points. We reconstruct pseudo-1982 profiles from the 2000 census marriage histories using birth-year and completed-age indices. Their differentials are 3.95 and 3.86 points; on identical respondents, the completed-minus-birth-year index difference is -0.08 points (wild $p = 0.341$). The birth-year-indexed pseudo-minus-actual estimate is -0.28 points, with a wild 95-percent interval of $[-2.45, 2.20]$. The point estimates align across differently selected retrospective and resident-stock estimands, although the interval permits meaningful differences and neither current-residence design identifies a program effect. In an exact point decomposition, education shares account for 0.59 points, or 15.5 percent of the comparable collapsed gap; the 3.21-point within-education component is imprecise under wild inference ($p = 0.118$). A negative but imprecise sex-ratio estimate provides little support for a positive male skew. Risk-window-aligned cohort comparisons likewise do not deliver a robust Third-Front-specific campaign shift. The result is a reproducible demographic regularity and an estimand lesson, not an average treatment effect of the Third Front.

Keywords: marriage timing; industrialization; work units; place-based policy; Third Front; China.
JEL codes: J12, J23, N35, O25, P23.

---

*We thank Patricia Funk, Eugenia Frezza, and seminar participants at USI and the 2026 GLO Global Conference for helpful comments.
†Università della Svizzera italiana, Via Giuseppe Buffi 13, Lugano; yi.yao@usi.ch.
‡Università della Svizzera italiana, Via Giuseppe Buffi 13, Lugano; tai.lo.yeung@usi.ch.

1 Introduction

Industrialization changes family formation through two margins. It changes the return to work, and it changes the institutions through which young adults obtain jobs, housing, and permission to move. Recent evidence emphasizes the labor-demand margin: when manufacturing jobs disappear in the United States, the marriage-market value of young men falls and marriage recedes (Autor, Dorn and Hanson, 2019), and trade shocks move marriage and fertility in developing economies (Braga, 2018). These mechanisms presume that labor demand reaches workers as wages and employment probabilities. Industrialization under central planning worked differently. A factory raised local productive capacity while placing workers inside an organization that assigned jobs, rationed housing, controlled residence, and supplied the paperwork needed to form a household.

This paper documents the demographic profile such places carry and then aligns the empirical objects needed to interpret it. In China's 1982 census, taken two years after the Third Front construction campaign wound down, prefectures with heavier industrial composition inside the designated provinces hold far more never-married young adults than the same comparison elsewhere: 4.2 percentage points more at ages 20–24 than at ages 30–34, per standard deviation of exposure. The profile looks like a marriage queue—an institution holding couples back in their twenties and releasing them—and a work-unit system that allocated housing by seniority makes that reading tempting. A cross-sectional age profile alone cannot tell whether the shape reflects marriage histories or resident composition. The 2000 census histories let us reconstruct marital status for the corresponding reported birth years. Birth-year and completed-age reconstructions both produce similar point estimates to the original, without supplying individual linkage or historical residence. A separate question is causal attribution: the young and old 1982 groups both passed through marriage risk during the campaign, so neither their cross-sectional contrast nor its retrospective counterpart is a treated-versus-pre-program comparison.

The distinction among three empirical objects organizes everything. A cross-sectional never-married share is a *stock* among current residents; a retrospective history describes marriage timing within an observed *cohort*; and a policy *effect* requires exposure assigned independently of subsequent residence and outcomes. The distinction is classical: age-specific proportions never married summarize a population at a date (Hajnal, 1953), while cohort interpretations of repeated cross sections require membership that is consistently recoverable across samples (Deaton, 1985; Ryder, 1965). Selective mobility can make repeated-cross-sectional profiles diverge sharply from longitudinal histories (Lubotsky, 2007). As an organizing accounting relationship rather than a new estimator, if g indexes groups defined by cohort and observed or unobserved characteristics, the resident stock at age a is

$$R_{pa} = \sum_{g} \Pr(g \mid p, a, \mathcal{R}) \, [1 - F_{gp}^{\mathcal{R}}(a)], \tag{1}$$

where $\mathcal{R}$ denotes selection into the observed resident population and $F_{gp}^{\mathcal{R}}(a)$ is the first-marriage distribution within that selected population. An industrial-place gradient in R_{pa} can arise from group-specific timing or from the resident-composition weights; a repeated cross section cannot separate them without additional assumptions or longitudinal residence information. We therefore place the stock beside a birth-year-indexed retrospective reconstruction and a separate campaign-cohort contrast. The first has a similar point estimate; the second answers a different campaign-timing question.

The data are the one-percent microdata of the 1982, 1990, and 2000 population censuses (Minnesota Population Center, 2020), aggregated to prefecture-by-single-age-by-sex cells on a province-validated crosswalk of 148 prefectures in 21 provinces, held fixed across censuses. Industrial composition comes from a 1980 heavy-industry share reconstructed from 253 million firm-registration records and, as a sharper alternative, from the plant-employment, investment, and designated-center measures in the replication files of Fan and Zou (2021). The design is an age-profile difference-in-differences-in-differences: within each prefecture, never-married rates at ages 20–24 and 25–29 are contrasted with rates at 30–34, and the question is whether that contrast steepens with industrial composition more inside Third Front provinces than outside, under prefecture fixed effects and province-by-age-by-sex effects.

Four findings organize the paper. First, the 1982 resident-stock differential is large and positive under the baseline analytic and wild procedures: 4.23 percentage points per standard deviation (province-clustered $p = 0.019$; restricted-null wild bootstrap over 21 provinces, $p = 0.039$; exact reassignment of the nine designation labels among the 13 interior provinces, $p = 0.059$). The net slope inside designated provinces is 4.82 points against 0.59 elsewhere, so the pattern is a young-age gradient of the designated interior, not the difference of two opposite slopes. Precision weakens materially with siting adjustment: predetermined controls reduce the differential to 2.24 points (wild $p = 0.099$), and the unweighted version has $p = 0.128$. We therefore report 2.2–4.2 points as a descriptive range rather than one invariant effect.

Second, age-index alignment narrows the gap between the stock and the histories. For the reported 1948–1962 birth years in the 2000 census, we reconstruct pseudo-1982 status using 1982−birth year. We then re-estimate both this index and recorded age in 2000 minus 18 on the 1,316,216 respondents common to their age-20–34 supports. The full birth-year and matched completed-age differentials are 3.95 and 3.86 points (wild $p = 0.015$ and 0.014), compared with 4.23 among actual 1982 residents. On identical respondents, the completed-minus-birth-year index difference is only −0.08 points (wild $p = 0.341$). Joint stacked models estimate pseudo-minus-actual differences of −0.28 and −0.37 points (wild $p = 0.797$ and 0.714). A marriage-year construction gives 3.75 points. Similar point estimates across differently selected resident populations do not identify a mechanism or establish equivalence, but they show that the original shape survives two transparent age-index translations into retrospective histories.

Third, age-index alignment is not program identification. The original 20–24 and 30–34 groups reached marriageable ages largely within the same 1964–1980 build-out, so their difference cannot isolate the campaign. In broader build-out-versus-pre-program histories, the Third-Front-versus-other differential is imprecise throughout ages 21–28. The net slope shift inside designated provinces is more informative about the raw pattern: it reaches roughly −1.5 points by age 23 and then catches up by 25. That early deficit is pointwise but not family-wise robust across eight thresholds. Threshold-specific comparisons whose entire marriage-risk windows lie either before 1964 or within 1964–1980 lead to the same disciplined conclusion: this current-residence design does not recover a robust Third-Front-specific campaign effect.

Fourth, observed education composition is a minority component in the point decomposition, not the explanation by default. A symmetric Kitagawa decomposition of the comparable collapsed young–old gap assigns 0.59 of 3.80 points (15.5 percent) to different education shares and 3.21 points (84.5 percent) to marriage-rate differences within education groups (Kitagawa, 1955); the latter component has wild $p = 0.118$. Students and collective dwellings do not account for the profile, and an exact-support sex-ratio regression has a negative but imprecise point estimate, providing little support for a positive male skew. Institutional facts remain suggestive context—in the 1988 urban household survey, the adjusted industrial-versus-other public-housing gap is 11 points more positive in designated provinces (wild $p = 0.017$), and case evidence documents socially isolated factory communities (Marukawa, 2021)—but they do not identify a housing-queue mechanism. Sharper treatment measures also weaken under siting controls, and the railway design is inconclusive under weak-instrument-robust few-cluster inference (Anderson–Rubin wild $p = 0.366$).

The contribution is a policy-specific estimand alignment, not the general observation that stocks and cohorts differ and not a new estimate of the Third Front's average treatment effect. Historical place-based programs are routinely assessed with repeated censuses because censuses are often all that survives (Kline and Moretti, 2014). Here an apparent contradiction narrows when reported birth years and completed-age definitions are aligned: the retrospective profiles have point estimates close to the resident-stock gradient. Yet the difference intervals allow meaningful departures, and point-estimate agreement does not establish treatment because residence is measured after exposure and the cross-sectional age groups do not separate pre-program from build-out risk. Showing where measurement alignment succeeds and where program attribution still fails is the substantive result.

Within the Third Front literature, Fan and Zou (2021) study industrial siting and local development, and Fang, Liang and Qin (2026) study human-capital consequences with family outcomes downstream. Closest in question is an unpublished county-level factory-timing project by Gao, Zhang, and Zhou, publicly described as finding later marriage among exposed women.[1] We claim no priority on Third Front family effects and do not treat our prefecture-composition design as a substitute for factory timing.

The closest broader work also clarifies our object. Tian (2016) uses the same census waves to trace synthetic-cohort transitions to adulthood, while Song and Zheng (2016) studies a historical Chinese policy and marriage timing with exposure defined before the outcome. Place-based-policy research emphasizes that local outcomes combine treatment, sorting, and equilibrium adjustment (Busso, Gregory and Kline, 2013; Criscuolo et al., 2019; Glaeser and Gottlieb, 2008; Kline and Moretti, 2013; Gerritse, Wang and van Oort, 2026; Kline and Moretti, 2014); labor-demand studies of family formation use predetermined exposure or plausibly exogenous shocks (Autor, Dorn and Hanson, 2019; Braga, 2018); and Chinese marriage-market research shows how migration and education reshape local partner pools (Angrist, 2002; Dupuy, 2021; Xiong, 2023; Mu and Yeung, 2020; Fang et al., 2026; Rossi and Xiao, 2026). Our contribution is narrower: we ask whether an industrial gradient defined by current residence survives translation into an aligned cohort-timing object.

The design's limits are stated where they bind: the registry share is an endogenous reconstruction, current residence is not birthplace, and no dataset here contains plant

[1] The authors' public research page lists the project as under review, and an institutional seminar abstract describes county-by-factory variation from roughly 600 plants (Gao, Zhang and Zhou, n.d; School of Labor Economics, Capital University of Economics and Business, 2025). No manuscript, data, or replication package was publicly available to us as of 30 August 2026; we report only what those materials state and cannot assess implementation or inference. Because our histories assign exposure by residence in 2000, our results and their described design can coexist: a factory-level effect on childhood exposure need not appear in a current-residence prefecture contrast.

activation dates. Section 2 describes the campaign and work-unit institution. Section 3 describes the data and crosswalk audit. Section 4 defines the estimand ladder. Section 5 presents the resident profile and age-index bridge, Section 6 the campaign-cohort comparisons, and Section 7 the formal composition accounting and secondary evidence. Section 8 delimits identification, and Section 9 concludes.

2 The Third Front and Household Formation

The Third Front was launched in 1964 after Chinese leaders came to view the coastal industrial base as vulnerable to attack. The campaign moved defense and heavy industry into the Southwest, Northwest, and interior Central China, building rail, power, and urban infrastructure around it. Siting followed strategic and topographic criteria—dispersed, concealed, often mountainous locations away from existing cities—and investment slowed in the late 1970s, though many plants remained major employers afterward. For identification these features pull in opposite directions: siting responded less to market demand than a typical industrial expansion, but projects were not randomly placed within the interior; rail access, terrain, existing capacity, and administrative priorities all mattered. Fan and Zou (2021) address placement with distance to the pre-1962 railway network, conditional on the completed 1980 network and geography; we use their variables as a demanding robustness exercise without inheriting their exclusion restriction for marriage.

Urban industrial employment was organized through the *danwei*. A work unit was employer, welfare provider, and gatekeeper at once: it assigned work, administered wages, allocated housing, and—until the 2003 regulations—supplied the marital-status documentation required to register a marriage (Davis, 2014). The bundling admits mechanisms of both signs. Housing queues and seniority rules could delay marriage; stable jobs and subsidized housing could ease it; isolated compounds could narrow partner pools; administrative assignment could alter education and its opportunity cost of early marriage. The historical record makes isolation concrete: in Shanghai's Small Third Front in southern Anhui, factories imported inputs and employees from Shanghai and were socially separate from surrounding villages, with most employees marrying within the industrial community (Marukawa, 2021). A queue account carries a specific timing implication—more singlehood at young ages *and* later catch-up within the same cohorts—and observing only the first, cross-sectionally, is insufficient. The work-unit system also outlived the construction campaign: job assignment, public housing, and residence controls remained central through the 1980s and unwound gradually in the 1990s, which is why repeated censuses and the 2000 retrospective histories provide informative boundaries.

3 Data

Outcomes come from the one-percent samples of the 1982, 1990, and 2000 population censuses distributed by IPUMS International (Minnesota Population Center, 2020). The cross-sectional profiles retain ages 18–40 with valid age, sex, marital status, weight, and geography and aggregate to prefecture-by-single-age-by-sex cells with person weights. The outcome is an indicator for never married. The microdata also identify current students, collective dwellings, and education strata; the 1990 census records residence in the same minor administrative unit five years earlier, and the 2000 census the same major unit, a province-level stability measure we label literally. The 2000 census additionally reports age and year at first marriage for the ever married; combined with current status, this yields marriage-by-each-age outcomes among older respondents who were unmarried at twenty, with the two constructions agreeing within one year for 99.995 percent of ever-married respondents. The broad history analysis covers 3.22 million people; the 1948–1962 birth-year bridge maps 1.34 million 2000 respondents to the common geography. These data support synthetic-cohort descriptions (Tian, 2016), but current prefecture is not historical residence.

Because prefecture names are not unique across China, the census-to-prefecture crosswalk parses the province from each IPUMS geographic code and requires it to match the first two digits of the 1982 prefecture code; name-only matching would admit 13 false candidates in 1982 and 10 in 1990. The province-constrained match yields 188 prefectures in 1982 and 233 in 1990; intersecting the 1982–1990–2000 supports with a valid exposure leaves 154, and dropping provinces contributing a single prefecture leaves 148 in 21 provinces. All cross-census comparisons hold this geography fixed, and a validation script enforces crosswalk uniqueness, sample counts, and manifest freshness (Appendix A).

Broad exposure is the heavy-industry share—manufacturing plus mining—of cumulative registered entries in a prefecture through 1980, constructed from 253 million raw firm-registration records with founding years in an author-licensed MacroData extract downloaded in August 2026 (MacroData, 2026). The registration system was rebuilt around the reform era, so pre-1980 records are retrospective founding dates rather than a contemporaneous plant census, and measurement error need not be classical; the variable is best understood as a noisy record of inherited industrial composition. Exposure distributions overlap extensively across designation status (means 0.500 and 0.469; Appendix Figure 3). The overlap reduces extrapolation but validates neither linearity nor exogeneity. As sharper alternatives we use the Fan and Zou (2021) prefecture files: 1985 large- and medium-plant employment relative to 1982 employment, Third Front investment, the ten designated industrial centers, rail distances, and siting controls. Designation follows the historical program at the province level (Yunnan, Guizhou, Sichuan including Chongqing, Shaanxi, Gansu, Ningxia, Qinghai, Shanxi, Henan, Hubei, Hunan) and is absorbed by prefecture fixed effects. Qinghai and Ningxia contribute only singleton prefectures and drop from estimation, leaving nine designated provinces on the 148-prefecture support.

Two auxiliary sources measure institutions and provide an independent survey design. The 1988 urban China Household Income Project (CHIP) records whether the household head worked in industry and whether the dwelling was publicly owned, for 8,094 households in ten provinces, five designated (Griffin and Zhao, 2010). The 2010–2022 China Family Panel Studies (CFPS) waves provide retrospective first-marriage histories (Xie and Hu, 2014; Institute of Social Science Survey, Peking University, 2015); because the access-controlled project extract identifies survey-era city rather than pre-marriage location, its estimates are secondary (Section 7). City and respondent identifiers are not redistributable and are excluded from the public replication package.

4 Empirical Design

Let N_{pas} be the weighted never-married share in prefecture p, age a, sex s; let H_p be the 1980 heavy share standardized by its cross-prefecture standard deviation (0.137, computed on the 154-prefecture common support before dropping singleton provinces), and T_p Third Front designation. Estimation uses 148 prefectures in 21 provinces. The core specification is

$$N_{pas} = \alpha_p + \lambda_{q(p)as} + \sum_{g \in \{20\text{–}24,\, 25\text{–}29\}} \left[\beta_g H_p \mathbf{1}\{a \in g\} + \delta_g H_p T_p \mathbf{1}\{a \in g\}\right] + \varepsilon_{pas}, \qquad (2)$$

with ages 30–34 omitted, prefecture fixed effects α_p, and province-by-single-age-by-sex effects $\lambda_{q(p)as}$; cells are weighted by population and clustered by province, and stacked versions add census interactions. The headline coefficient $\delta_{20\text{–}24}$ asks whether the relationship between industrial composition and young singlehood bends more sharply at young ages inside designated provinces, relative to the same prefectures at 30–34. Prefecture effects absorb permanent local attributes, including designation itself; the saturated province-age-sex effects absorb every province-specific marriage-age profile. What survives as a threat is any prefecture characteristic that covaries with H_p and tilts the age profile differentially inside designated provinces—education is the leading example, and the results treat it as such. With 21 clusters, analytic p-values are supplemented by restricted-null wild cluster bootstraps (4,999 draws) and by exact reassignment of the nine designation labels among the 13 interior provinces; historical assignment was not uniform over that set, so the exact tail areas are design diagnostics, not randomization inference.

The age-index bridge first asks a measurement question. For each 2000 respondent with reported birth year 1948–1962, the primary index is $a^{BY} = 1982-$birth year; we reconstruct whether first marriage had occurred by a^{BY}. Actual 1982 cells instead use completed age, and birth months are unavailable, so this is a near alignment rather than exact linkage. The completed-age sensitivity uses a^{CA} =recorded completed age in 2000 minus 18. The unrestricted age-20–34 boundary samples under a^{BY} and a^{CA} contain 1,339,802 and 1,327,590 respondents, with 1,316,216 in both. We re-estimate both indices on those identical respondents so their direct difference isolates indexing rather than boundary composition. Within the full birth-year sample, the two indices differ by one year for 17.1 percent of records; different census months mean that neither exactly recovers completed age on the 1982 census date. We re-estimate Equation (2) on each pseudo-1982 profile using residence in 2000 and stack actual and pseudo cells with source-specific prefecture and source-specific province-by-age-by-sex effects. Pseudo-minus-actual interactions, their covariance, and wild tests are therefore estimated jointly over the same 21 provinces. Age at first marriage is primary; marriage year provides a third calendar-age sensitivity.

The campaign comparison asks a different question and reports three slopes. Cohorts born 1944–1960 reached age twenty during the build-out, while cohorts born 1934–1943 did so before it. For marriage by each age $a = 21, \ldots, 28$ among people unmarried at twenty, a model with prefecture and province-by-single-birth-year-by-sex effects reports the slope shift outside designated provinces, the additional Third-Front differential, and their sum—the net shift inside designated provinces. We use restricted-null wild inference for both the differential and the net shift and adjust each eight-threshold family by Holm's method. Because these broad cohort labels still permit marriage-risk windows to straddle 1964, a stricter diagnostic changes the cohorts at each threshold: the full ages-21-to-a risk window must end by 1963 or lie wholly within 1964–1980, and the nearest pre-program window contains the same number

Table 1: The estimand ladder

Empirical object	Residence and cohort comparison	What it establishes	What it cannot establish
Actual 1982 stock	Current residents; ages 20–24 versus 30–34	Resident-population age gradient in 1982	Individual histories or a campaign effect
Pseudo-1982 history	2000 survivors/residents; reported birth-year or completed-age index	Whether the gradient survives a retrospective near-alignment	Exact 1982 completed age, the 1982 resident population, or causality
Calendar-clean history	2000 survivors/residents; risk wholly before or within 1964–1980	Cohort slope shifts under non-overlapping risk windows	Effects on movers or a factory-arrival treatment effect
Factory-timing design	Birthplace or pre-marriage residence linked to activation dates	Would permit a cohort-by-project contrast under explicit placement and timing assumptions	Not available in the present data

Notes. The table separates measurement alignment from treatment identification. Agreement between the first two rows is substantive evidence about the profile, but it does not promote either row into the fourth.

of birth cohorts as the build-out window. This removes calendar overlap at the cost of older, survivor-selected comparison cohorts. Every history design assigns exposure by residence in 2000, not by birthplace or residence when marriage risk began.

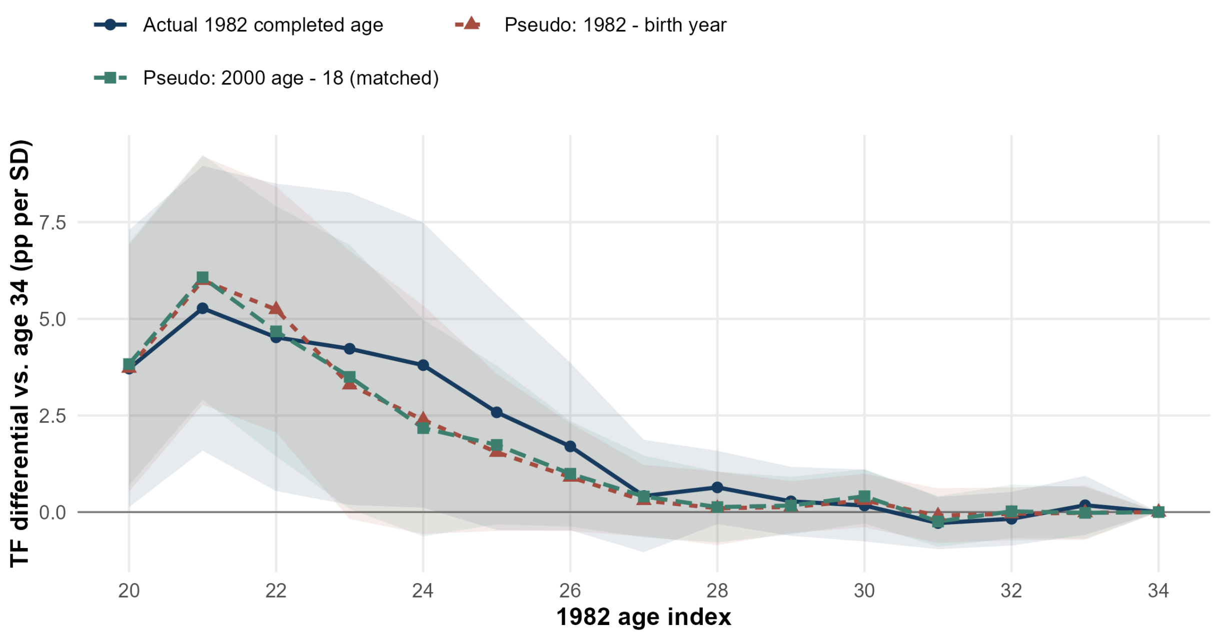
Table 1 summarizes which population and timing comparison each design recovers.

Finally, for direct capacity we reproduce the Fan and Zou (2021) railway design inside their hinterland: plant employment interacted with the young-age indicator, instrumented by 1962 rail distance interacted with the same indicator, conditional on their siting controls. The usable overlap is 22 prefectures, so we let the reduced-form wild-cluster Anderson–Rubin test discipline the conclusion.

5 Resident Stocks and the Age-Index Bridge

Figure 1 puts the paper's central near-alignment in one place. The solid profile uses actual 1982 residents. Replacing the age bins in Equation (2) with single-year interactions (age 34 reference), the differential peaks near 5.3 percentage points per standard deviation at age 21, remains elevated through 24, and approaches zero around age 30. The retrospective profiles reconstruct never-married status using the birth-year and completed-age indices defined above. Their point profiles are similar to the actual stock despite different survival and residence selection. This is a distinct check from recurrence in another cross section, but it remains weaker than individual linkage or an equivalence result; it neither recovers exact 1982 completed age nor reveals where the 2000 respondents lived in 1982.

Figure 1: Resident stocks and retrospective histories under alternative age indices



Notes. The actual series uses never-married stocks among 1982 residents. Retrospective series reconstruct pseudo-1982 status from age at first marriage among 2000 survivors/residents, indexing age either as 1982 minus reported birth year or as 2000 completed age minus 18. The completed-age series uses the 1,316,216 respondents whose two indices both lie from 20 to 34; the full birth-year series retains all 1,339,802 reported 1948–1962 births. Within that full sample, the indices differ by one year for 17.1 percent of records. Coefficients are age-specific heavy-share-by-Third-Front differentials relative to age 34, scaled by the common-support standard deviation. Models contain prefecture and province-by-age-by-sex effects, use population weights, and cluster by province. Pseudo profiles assign exposure by residence in 2000 and do not reconstruct the 1982 resident population.

Table 2 contains the grouped estimates and the full inference. The 1982 baseline differential is 4.23 points (s.e. 1.66) at 20–24 and 1.14 (0.66) at 25–29; the net slope inside designated provinces is 4.82 points against 0.59 outside, so the differential is primarily a young-age gradient of the designated interior. Wild-bootstrap inference over the 21 provinces gives $p = 0.039$ at 20–24 and $p = 0.105$ at 25–29, and no single province drives the estimate: leave-one-province-out estimates span 3.54 to 5.12 points with 20 of 21 intervals excluding zero

(Appendix Figure 4). The diagnostics do not reject linearity (median-knot spline $p = 0.67$), but the nonmonotone quartile pattern cautions against treating the slope as a structural dose response: the top-versus-bottom contrast is 8.99 points ($p = 0.039$), while the quartile coefficients are not jointly distinguishable ($p = 0.19$; Appendix Table 6).

Table 2: The Third Front differential in the census marriage-age profile

	Ages 20–24		Ages 25–29		Net TF slope, 20–24	
Panel A. 1982 census						
Baseline	+4.23**	(1.66)	+1.14	(0.66)	+4.82***	(0.86)
Predetermined siting controls	+2.24*	(1.15)	+0.18	(0.49)	+2.16***	(0.74)
Siting controls, unweighted	+2.51	(1.58)	—		—	
Nonstudents, household dwellings	+4.38**	(1.69)	+0.91	(0.62)	+4.91***	(0.87)
Education-stratum fixed effects	+2.27	(1.43)	+0.86	(0.52)	+2.69***	(0.66)
Panel B. Later censuses, same geography						
1990 baseline	+2.33*	(1.26)	+0.31	(0.62)	+2.60**	(0.92)
1990, same minor unit 1985	+2.27*	(1.31)	+0.42	(0.58)	+2.50**	(0.97)
2000 baseline	+3.15***	(0.73)	+1.42***	(0.46)	+3.07***	(0.62)
2000, nonstudent, same major unit 1995	+2.74***	(0.84)	+1.28***	(0.42)	+2.78***	(0.81)
2000, education-stratum fixed effects	+2.28***	(0.64)	+1.11***	(0.34)	+1.83***	(0.54)
Stacked fade, 1990 minus 1982	−1.90*	(0.97)	−0.83*	(0.43)	—	
Stacked fade, 2000 minus 1982	−1.08	(1.51)	+0.28	(0.49)	—	

Notes. Equation (2) on the common 148-prefecture, 21-province geography; cells weighted by population; standard errors clustered by province. Entries are percentage points per cross-prefecture standard deviation (0.137) of the 1980 heavy-industry share; age bins are contrasts against ages 30–34 within prefecture. The net Third Front slope is the sum of the ordinary and interaction slopes. Restricted-null wild cluster bootstrap over the 21 provinces (4,999 draws): baseline 1982 differential $p = 0.039$, siting-control differential $p = 0.099$. Exact reassignment of the nine designation labels among the 13 interior provinces: baseline $p = 0.059$, siting-controlled $p = 0.249$. Leave-one-province-out baseline estimates span 3.54 to 5.12 points, with 20 of 21 conventional 95-percent intervals excluding zero. Siting controls interact pre-campaign investment, 1962 rail distance, terrain, elevation, distance to the provincial capital, 1964 urbanization and density, and 1936 employment with the age bins. The completed 1980 rail network is post-campaign and deliberately excluded. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Sensitivity to siting is the estimate's most important property after its existence, and we give it equal billing. Interacting predetermined siting covariates—pre-campaign investment, 1962 rail distance, terrain, elevation, distance to the provincial capital, 1964 urbanization and density, and 1936 employment—with the age bins reduces the differential to 2.24 points (s.e. 1.15; wild $p = 0.099$), and the unweighted version is 2.51 ($p = 0.128$). The completed 1980 rail network is post-campaign and therefore excluded from this control set. Under exact designation reassignment the controlled estimate's tail area is 0.249, against 0.059 for the baseline. The defensible summary is a descriptive range of roughly 2.2 to 4.2 points whose magnitude and precision depend on how much of the program's siting geography one holds fixed. Converting that stock magnitude into months of delay would require stronger residence and cohort assumptions than the data supply.

Panel B of Table 2 shows recurrence on the identical geography. The 20–24 differential is 2.33 points in 1990 and 3.15 in 2000; a stacked model estimates the 2000-minus-1982 change at -1.08 ($p = 0.48$), while the 1990-minus-1982 dip is -1.90 (wild $p = 0.114$). These repetitions establish a recurring resident-age profile, not a persistent cohort effect: age and cohort determine census year, so the repeated cross sections cannot separately identify all three. The 2000 restriction to nonstudents in the same major unit five years earlier gives 2.74 points, and the 1990 same-minor-unit subsample gives 2.27 (s.e. 1.31); these five-year restrictions do not address lifetime sorting.

Table 3 aggregates the profiles in Figure 1. The actual, full-sample birth-year, and matched-sample completed-age 20–24 differentials are 4.23, 3.95, and 3.86 points. Re-estimating the birth-year index on the same matched records gives 3.95 points, and the direct completed-minus-birth-year difference is -0.08 points (s.e. 0.08; wild $p = 0.341$). In joint stacked models, the two pseudo-minus-actual estimates are -0.28 points (s.e. 1.01; wild $p = 0.797$; wild interval $[-2.45, +2.20]$) and -0.37 points (s.e. 1.00; wild $p = 0.714$; interval $[-2.41, +2.11]$). Reconstructing timing from marriage year gives 3.75 points and a difference of -0.48 (wild $p = 0.639$). These are comparisons of differently selected estimands, not equivalence tests and not evidence that migration selection is absent; marriage-related relocation could itself make current 2000 residence line up with past marriage histories.

Table 3: Age-index bridge between the 1982 resident stock and retrospective histories

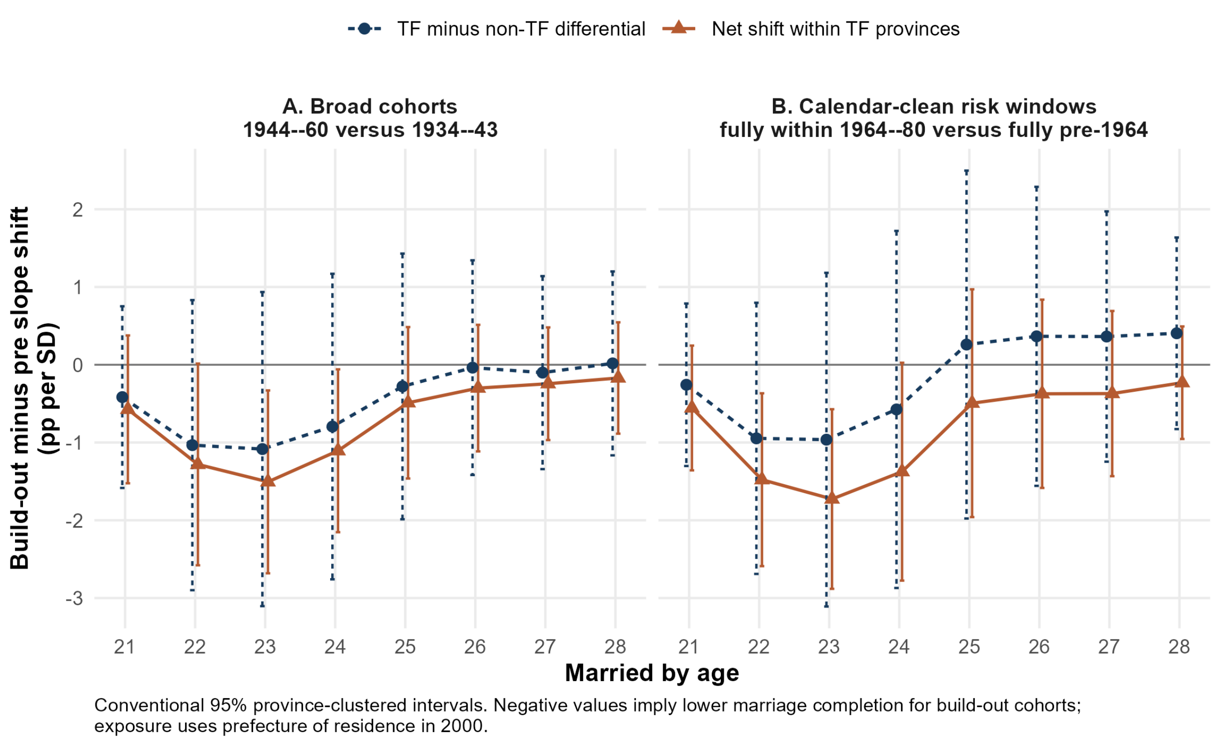
	Ages 20–24 versus 30–34			Ages 25–29 versus 30–34		
	Estimate	(s.e.)	Wild p	Estimate	(s.e.)	Wild p
Panel A. Third Front differential by estimand						
Actual 1982 resident stock	+4.23	(1.66)	0.039	+1.14	(0.66)	0.105
Pseudo: 1982 minus birth year, marriage age	+3.95	(1.37)	0.015	+0.56	(0.41)	0.205
Pseudo: completed-age index, matched records	+3.86	(1.36)	0.014	+0.65	(0.42)	0.150
Pseudo: 1982 minus birth year, marriage year	+3.75	(1.42)	0.025	+0.72	(0.49)	0.185
Panel B. Pseudo minus actual						
Birth-year index, marriage age	−0.28	(1.01)	0.797	−0.58	(0.39)	0.145
Completed-age index, matched records	−0.37	(1.00)	0.714	−0.49	(0.39)	0.205
Birth-year index, marriage year	−0.48	(1.00)	0.639	−0.41	(0.37)	0.272
Panel C. Index difference on identical respondents						
Completed-age minus birth-year index	−0.08	(0.08)	0.341	+0.09	(0.05)	0.114

Notes. Entries are percentage points per cross-prefecture standard deviation (0.137) of the 1980 heavy-industry share. All models use the common 148-prefecture, 21-province support, prefecture fixed effects, province-by-single-year-age-by-sex effects, population weights, and province clustering. The actual estimand is the never-married stock among 1982 residents. The primary pseudo profile uses the exact reported 1948–1962 birth cohorts and indexes age as 1982 minus birth year. The completed-age sensitivity uses the 1,316,216 respondents whose birth-year and completed-age indices both lie from 20 to 34. Panel C re-estimates both indices on those identical respondents, isolating the index change from boundary composition. Within the full birth-year sample, 17.1 percent differ by one year under the two indices. Neither observes birth month or exact completed age on the 1982 census date. Both use 2000 first-marriage histories, assign exposure by residence in 2000, and condition on survival and residence in China through 2000. Wild p-values use 4,999 restricted-null Rademacher draws over provinces.

6 Campaign Cohorts and Risk Windows

The aligned retrospective profile answers whether the original age shape is visible for the corresponding reported birth years; it does not compare cohorts whose marriage risk fell before and during the campaign. Figure 2 and Table 4 turn to that second question. They report the change in the heavy-industry slope from pre-program to build-out cohorts outside designated provinces, the additional Third-Front differential, and their sum—the net slope change inside designated provinces. A negative marriage-by-age estimate means that industrial intensity is associated with a larger marriage deficit for build-out cohorts.

Figure 2: Broad and calendar-clean cohort timing across fixed-age thresholds



Notes. The broad comparison uses births 1944–1960 versus 1934–1943. The calendar-clean comparison changes the birth windows at each threshold so the full ages-21-to-threshold risk period falls either before 1964 or within 1964–1980; the two windows contain equal numbers of birth cohorts and sit next to the policy boundary. Outcomes are marriage by the indicated age among people unmarried at 20. Models include prefecture fixed effects and province-by-single-birth-year-by-sex effects, weight by population, and cluster by province. Error bars are conventional 95-percent intervals; Table 4 reports restricted-null wild and Holm-adjusted p-values. Current prefecture in 2000 assigns exposure.

Table 4: Cohort timing in the 2000 census: broad and calendar-clean comparisons

Threshold	Ordinary non-TF	TF minus non-TF differential			Net shift within TF		
	Estimate (s.e.)	Estimate (s.e.)	Wild p	Holm p	Estimate (s.e.)	Wild p	Holm p
Panel A. Broad cohorts: births 1944–60 versus 1934–43							
Married by 21	−0.16 (0.33)	−0.42 (0.56)	0.515	1.000	−0.57 (0.46)	0.290	1.000
Married by 22	−0.25 (0.64)	−1.03 (0.89)	0.310	1.000	−1.28 (0.62)	0.078	0.475
Married by 23	−0.42 (0.79)	−1.08 (0.97)	0.305	1.000	−1.51 (0.56)	0.039	0.315
Married by 24	−0.31 (0.80)	−0.79 (0.94)	0.444	1.000	−1.11 (0.50)	0.068	0.475
Married by 25	−0.21 (0.67)	−0.28 (0.82)	0.741	1.000	−0.49 (0.47)	0.336	1.000
Married by 26	−0.26 (0.53)	−0.04 (0.66)	0.958	1.000	−0.30 (0.39)	0.462	1.000
Married by 27	−0.14 (0.48)	−0.10 (0.59)	0.865	1.000	−0.24 (0.35)	0.527	1.000
Married by 28	−0.19 (0.45)	+0.02 (0.57)	0.980	1.000	−0.17 (0.34)	0.627	1.000
Panel B. Calendar-clean risk windows: fully within 1964–80 versus fully pre-1964							
By 21 [pre 1926–1942; build 1943–1959]	−0.30 (0.32)	−0.26 (0.50)	0.658	1.000	−0.56 (0.38)	0.223	1.000
By 22 [pre 1926–1941; build 1943–1958]	−0.53 (0.64)	−0.95 (0.84)	0.300	1.000	−1.48 (0.53)	0.015	0.118
By 23 [pre 1926–1940; build 1943–1957]	−0.76 (0.87)	−0.96 (1.03)	0.386	1.000	−1.73 (0.55)	0.027	0.188
By 24 [pre 1926–1939; build 1943–1956]	−0.80 (0.87)	−0.57 (1.10)	0.639	1.000	−1.38 (0.67)	0.090	0.543
By 25 [pre 1926–1938; build 1943–1955]	−0.75 (0.81)	+0.26 (1.07)	0.824	1.000	−0.49 (0.70)	0.517	1.000
By 26 [pre 1926–1937; build 1943–1954]	−0.74 (0.72)	+0.37 (0.92)	0.707	1.000	−0.37 (0.58)	0.583	1.000
By 27 [pre 1926–1936; build 1943–1953]	−0.73 (0.58)	+0.36 (0.77)	0.654	1.000	−0.37 (0.51)	0.499	1.000
By 28 [pre 1926–1935; build 1943–1952]	−0.64 (0.48)	+0.40 (0.59)	0.519	1.000	−0.23 (0.35)	0.533	1.000

Notes. Entries are build-out-minus-pre changes in the heavy-share slope, in percentage points per common-prefecture standard deviation. Ordinary non-TF is the slope change outside Third Front provinces; the differential is the additional change inside Third Front provinces; net within TF is their sum. Parentheses contain conventional standard errors clustered by province. Wild p-values use 4,999 restricted-null wild-cluster draws over 21 provinces; the net null is imposed directly after reparameterizing the primitive coefficients. Holm p-values control the family-wise error rate across the eight thresholds, separately for the differential and net estimands within each panel. Panel A compares births 1944–60 with 1934–43, so threshold-specific marriage calendars can overlap the 1964 and 1980 boundaries. Panel B removes all boundary-straddling risk windows: for threshold a, pre-program births run from 1926 through $1963 - a$ and build-out births from 1943 through $1980 - a$, using the equal-sized cohorts closest to the boundary. This calendar separation comes at the cost of older, more survivor-selected pre-program cohorts. Outcomes are marriage by the indicated age among people unmarried at twenty. Models include prefecture fixed effects and province-by-single-birth-year-by-sex effects, use census person weights, and assign exposure by prefecture of residence in 2000.

The broad comparison reveals a short-delay-and-catch-up pattern. The net slope inside designated provinces is -0.57 points by age 21, -1.28 by 22, -1.51 by 23, and -1.11 by 24; it narrows to -0.49 by 25 and -0.17 by 28. The by-23 net estimate has a pointwise restricted-null wild $p = 0.039$, but its Holm family-wise p across the eight thresholds is 0.315; no threshold survives family-wise adjustment. More importantly for a Third-Front-specific claim, the additional differential ranges from -1.08 to $+0.02$ points, its raw wild p-values are all at least 0.305, and every Holm-adjusted p-value is one. The data therefore contain an early marriage deficit correlated with industrial intensity inside the designated region, but not a distinct or family-wise-robust campaign differential.

The non-overlapping risk windows sharpen rather than overturn that reading. Across ages 21–28, the calendar-clean net estimates are -0.56, -1.48, -1.73, -1.38, -0.49, -0.37, -0.37, and -0.23 points. The raw wild p-values at 22 and 23 are 0.015 and 0.027, but become 0.118 and 0.188 after Holm adjustment. The corresponding Third-Front differentials never approach family-wise significance. The birth-year diagnostic in Appendix Figure 6 likewise shows no discrete differential break at the campaign boundary. This is suggestive evidence of an early industrial-place deficit followed by catch-up, not an identified effect of Third Front construction.

Three qualifications are binding. First, neither $p > 0.05$ nor a Holm correction establishes equivalence or rules out a modest, short-lived effect. Second, the clean pre-program windows include survivors born as early as 1926, increasing the scope for differential mortality and selection. Third, all histories assign exposure by prefecture of residence in 2000. A factory-level effect on people observed where they lived before exposure—including the kind of childhood-exposure design publicly described by Gao, Zhang, and Zhou—could escape this comparison. The history evidence disciplines the program claim; it does not refute a treatment effect that the available residence measure cannot recover.

Table 5: Education standardization and the exact-support sex-ratio diagnostic

Outcome	Differential (SE)	Wild p	Within-TF (SE)	Wild p
Panel A. Exact symmetric Kitagawa decomposition, 1982				
Observed young-old gap	3.80 (1.81)	0.078	4.38 (0.86)	0.005
Education-composition component	0.59 (0.18)	< 0.001	0.70 (0.15)	0.004
Common-weight within-education component	3.21 (1.74)	0.118	3.68 (0.81)	0.005
Panel B. Male share among never-married residents, 1982				
Ages 20–34	-1.47 (1.14)	0.235	-2.18 (0.68)	0.052
Ages 20–29	-1.40 (1.11)	0.231	-2.08 (0.69)	0.047

Notes. Differential entries are the heavy-share slope inside Third Front provinces minus the slope elsewhere; within-TF entries are the net heavy-share slopes inside Third Front provinces. Both are per a one-standard-deviation increase in the 1980 heavy-industry share (SD = 0.137). Panel A is in percentage points of the ages 20–24 minus ages 30–34 never-married gap. Education is below-secondary versus secondary-plus; the composition and common-weight within-education components add exactly to the observed gap at every prefecture-sex observation. Regressions contain province-by-sex fixed effects and use the harmonic mean of young and old population. This is a descriptive standardization/accounting exercise for the collapsed age-bin gap, not a causal mediation analysis and not the paper's single-age DDD. Panel B is in percentage points of the male share among never-married residents, contains province fixed effects, and weights by the number never married. All specifications use 148 province-safe prefecture identifiers in 21 provinces and cluster by province. Wild p-values use a 4,999-draw restricted-null Rademacher bootstrap.

7 Composition, Institutions, and Secondary Designs

The birth-year-indexed similarity does not rule out sorting, so we next ask whether specific observed composition margins account for the pattern. Secondary-schooling shares have a 4.33-point young–old differential in the core 1982 design (Appendix Table 7). Merely adding education fixed effects reduces the marriage differential from 4.23 to 2.27 points, but coefficient attenuation across differently weighted fixed-effect specifications is not a decomposition. Table 5 instead applies the exact symmetric decomposition of Kitagawa (1955) to prefecture-by-sex young–old gaps. With two exhaustive education groups, the 3.80-point collapsed-gap differential separates exactly into 0.59 points from education shares and 3.21 points from marriage-rate differences within education. In this point decomposition education composition is 15.5 percent, not an explanation for half of the profile and not a causal mediator. The 84.5-percent within-education point component remains imprecise under wild inference ($p = 0.118$); it may contain institutional timing, unobserved composition within broad schooling groups, or both.

The remaining composition checks probe, but cannot eliminate, particular margins. Removing students and collective dwellings leaves the main marriage differential at 4.38 points. The profile appears for women (4.67 points) and men (3.79). On the audited 148-prefecture support, the male share among never-married residents ages 20–34 has a negative but imprecise -1.47-point differential (s.e. 1.14; wild $p = 0.235$), providing little support for the positive skew implied by a simple male-inflow squeeze. The interval remains too wide to rule out small positive shifts, and aggregate sex ratios cannot detect partner-pool segmentation within work-unit compounds. Five-year migration measures are likewise incomplete: recent-stayer restrictions leave the marriage profile close to baseline, but observe neither childhood location nor moves around first marriage. These results give little support to a simple sex-ratio mechanism (Angrist, 2002; Xiong, 2023; Mu and Yeung, 2020) without identifying the residual.

Institutional measurements explain why the setting invites a rationing interpretation. In the 1988 CHIP urban sample, industrial households were *less* dependent on public housing than their neighbors outside designated provinces (78.4 versus 85.2 percent) and *more* dependent inside them (90.2 versus 87.0); these raw cell means imply a 10.1-point difference-in-differences. The adjusted province-fixed-effect specification is 11.0 points (s.e. 3.6), with a ten-cluster wild-bootstrap $p = 0.017$ and an exact reallocation $p = 0.012$ (Appendix Table 10). This is directionally consistent with a work-unit housing channel, measured in the institution's own era. It remains context rather than a mechanism test: CHIP records no waiting list, allocation date, rejected certificate, or marriage outcome, while the census histories assign exposure only by later residence.

Secondary designs do not change that evidence ranking. The CFPS yields Third-Front-versus-other slope differentials of -1.7 points by 25 and -2.0 by 28, but the imprecisely estimated net slopes inside designated cities are only about -0.3 and -0.6; location is the survey-era city and the differential weakens under province-by-cohort effects (Appendix C). Project-capacity associations are sensitive to support, direct census investment and capacity profiles fall from 1.61 and 2.34 points to 0.46 and 0.92 under siting controls, and the designated-center coefficients are descriptive with only seven effective provinces. The railway IV's conventional estimate is positive, but its weak-instrument-robust wild Anderson–Rubin p-value is 0.366 on a 22-prefecture overlap (Appendix Tables 8 and 9). We do not use any of these secondary designs to select a mechanism.

The evidence hierarchy is consequently narrow. The retrospective reconstruction shows that the profile survives a birth-year-indexed history measure; it does not eliminate resident selection. The student, dwelling, education, and sex-ratio exercises probe only particular observed margins, with the within-education component itself imprecise under wild inference. Public-housing dependence and within-compound matching make a work-unit channel plausible, but neither is linked to individual waiting time or marriage. Lifetime sorting and marriage-related migration remain possible because historical residence is unobserved. Finally, the calendar-clean cohort design does not supply a robust Third-Front-specific treatment contrast. We leave the residual unassigned rather than choosing among mechanisms the data cannot separate.

8 What the Evidence Identifies

Three findings are defensible. First, Equation (2) estimates a Third-Front-versus-other differential in the young–old resident gradient: 4.23 points at baseline and 2.24 with siting controls on the maintained linear exposure scale. Second, a birth-year-indexed profile built from completed first-marriage histories has a similar point estimate, so the cross-sectional shape survives one retrospective-history reconstruction, subject to a wide difference interval. Third, education composition accounts for a measured minority of a comparable collapsed point estimate; student residence, collective dwellings, and the sex-ratio diagnostic provide little support for those specific accounts without excluding selection on other margins. None is an average treatment effect. The registry share is endogenous; the siting-control estimate and exact designation exercise show consequential geographic sensitivity. Actual and pseudo profiles select residents at different dates, and current residence can be an outcome of marriage. The histories do not observe birthplace or pre-marriage location, while the project files contain no activation dates. Thus the positive profile is not by itself evidence of a Third Front effect, a housing mechanism, or a welfare loss.

The converse limit matters just as much. A weak Third-Front-specific cohort differential does not establish equivalence and cannot rule out a factory-level effect on people observed where they lived before exposure. Aggregation, post-marriage migration, survival to 2000, and misassignment by current residence can attenuate or redirect a true effect. The calendar-clean comparison shows what this design can recover under non-overlapping risk windows; it does not substitute for project timing. This is why the present results can coexist with the factory-arrival effects described publicly by Gao, Zhang, and Zhou without either study refuting the other.

The most useful data upgrade is specific. A county-level roster of project approval, construction, and production dates, linked to birthplace or pre-marriage residence, would permit a cohort-by-county event study with pretrends and placement-appropriate inference; causal interpretation would still require evidence or assumptions addressing endogenous placement and timing. Restricted census birthplace codes would assign exposure before migration, while personnel or housing-allocation archives for even a few compounds would test the queue mechanism at the individual level.

9 Conclusion

Industrial Third Front prefectures exhibit a large young-singlehood differential: 4.23 percentage points per exposure standard deviation in the baseline and 2.24 with predetermined siting controls. That pattern is not confined to one cross section. Pseudo-1982 profiles reconstructed from later marriage histories using birth-year and matched completed-age indices are 3.95

and 3.86 points; changing the index on identical respondents moves the differential by only -0.08 points. The birth-year-indexed difference from the actual profile is -0.28 points, but its wild interval of $[-2.45, 2.20]$ is not an equivalence result. Nor does coarse education composition dominate the point estimate: exact standardization assigns about 16 percent of the comparable collapsed gap to education shares and 84 percent to within-education rate differences, with the latter imprecisely estimated under wild inference. The evidence supports a reproducible demographic pattern among the residents observed; it does not reveal why those residents are there or what the campaign caused.

The distinction is the paper's contribution. Cohort labels that look natural can answer different questions because the ages being contrasted passed through overlapping policy calendars. Once the history analysis reports outside, differential, and net Third Front slope shifts, the data allow a modest early deficit and catch-up inside designated provinces but do not yield a family-wise-robust Third-Front-specific campaign contrast. For the evaluation of historical place-based policy, the lesson is practical: align birth-year and completed-age definitions and the policy risk windows first, then ask whether exposure was assigned before migration. A persistent resident-age gradient can survive retrospective reconstruction and still fall short of a causal program effect.

What remains open is the question our data cannot reach: whether factory arrivals changed family formation for the specific people exposed to them, at the county and cohort level where the campaign operated. That answer requires project timing and pre-treatment residence, and we have specified the files that would provide both.

References

Angrist, Josh. 2002. "How do sex ratios affect marriage and labor markets? Evidence from America's second generation." *The Quarterly Journal of Economics*, 117(3): 997–1038.

Autor, David, David Dorn, and Gordon Hanson. 2019. "When work disappears: Manufacturing decline and the falling marriage market value of young men." *American Economic Review: Insights*, 1(2): 161–178.

Braga, Breno. 2018. "The effects of trade exposure on marriage and fertility choices: evidence from Brazil." IZA Discussion Paper 11875.

Busso, Matias, Jesse Gregory, and Patrick Kline. 2013. "Assessing the incidence and efficiency of a prominent place based policy." *American Economic Review*, 103(2): 897–947.

Criscuolo, Chiara, Ralf Martin, Henry G Overman, and John Van Reenen. 2019. "Some causal effects of an industrial policy." *American Economic Review*, 109(1): 48–85.

Davis, Deborah S. 2014. "Privatization of Marriage in Post-Socialist China." *Modern China*, 40(6): 551–577.

Deaton, Angus. 1985. "Panel Data from Time Series of Cross-Sections." *Journal of Econometrics*, 30(1–2): 109–126.

Dupuy, Arnaud. 2021. "Migration in China: To work or to wed?" *Journal of Applied Econometrics*, 36(4): 393–415.

Fang, Guanghua, Yinhe Liang, and Fan Qin. 2026. "The Long-Term Impact of Industrialization Campaigns on Human Capital Accumulation: Evidence from the "Construction of Third Front" in China." *Journal of Comparative Economics*, 54(2): 467–483.

Fang, Min, Zibin Huang, Yushi Wang, and Yu Yang. 2026. "Sex and the City: Demographics, Spatial Sorting, and the Marriage Market." *International Economic Review*.

Fan, Jingting, and Ben Zou. 2021. "Industrialization from scratch: The "Construction of Third Front" and local economic development in China's hinterland." *Journal of Development Economics*, 152: 102698.

Gao, Ming, Qingping Zhang, and Yi Zhou. n.d. "How Lives Change When Factories Arrive." *Contemporaneous research project listed as under review*, Public status verified August 30, 2026 at `https://yizhou-demog.github.io/research/`; no public manuscript was available to the authors.

Gerritse, Michiel, Zhiling Wang, and Frank van Oort. 2026. "Industrial transfer policy in China: Migration and development." *Journal of Urban Economics*, 151: 103815.

Glaeser, Edward L, and Joshua D Gottlieb. 2008. "The economics of place-making policies." National Bureau of Economic Research Working Paper.

Griffin, Keith, and Renwei Zhao. 2010. "Chinese Household Income Project, 1988 [dataset]." *ICPSR 9836, version 2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research*, Distributed July 6, 2010; DOI: `https://doi.org/10.3886/ICPSR09836.v2`.

Hajnal, John. 1953. "Age at Marriage and Proportions Marrying." *Population Studies*, 7(2): 111–136.

Institute of Social Science Survey, Peking University. 2015. "China Family Panel Studies (CFPS), 2010–2022 Waves [dataset]." *Peking University Open Research Data*

Platform, living repository, Persistent DOI: https://doi.org/10.18170/DVN/45LCSO; accessed through the official CFPS data platform in August 2026.

Kitagawa, Evelyn M. 1955. "Components of a Difference Between Two Rates." *Journal of the American Statistical Association*, 50(272): 1168–1194.

Kline, Patrick, and Enrico Moretti. 2013. "Place based policies with unemployment." *American Economic Review*, 103(3): 238–243.

Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics*, 6: 629–662.

Lubotsky, Darren. 2007. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy*, 115(5): 820–867.

MacroData. 2026. "Chinese Business-Registration Microdata [dataset]." *Author-licensed commercial extract*, Raw prefecture-address files downloaded August 2026; provider documentation at https://www.macrodatas.cn/.

Marukawa, Tomoo. 2021. "Isolated Societies under China's Planned Economy: Life and Production at Shanghai's "Small Third Front"." *Asian Studies*, 67(2): 21–37.

Minnesota Population Center. 2020. "Integrated Public Use Microdata Series, International: Version 7.3 [dataset]." Dataset DOI: https://doi.org/10.18128/D020.V7.3.

Mu, Zheng, and Wei-Jun Jean Yeung. 2020. "Internal Migration, Marriage Timing and Assortative Mating: A Mixed-Method Study in China." *Journal of Ethnic and Migration Studies*, 46(14): 2914–2936.

Rossi, Pauline, and Yun Xiao. 2026. "Left over or opting out? Squeeze, mismatch and surplus in Chinese marriage markets." *Journal of Development Economics*, 179: 103629.

Ryder, Norman B. 1965. "The Cohort as a Concept in the Study of Social Change." *American Sociological Review*, 30(6): 843–861.

School of Labor Economics, Capital University of Economics and Business. 2025. "Labor Economics Seminar No. 26: How Lives Change When Factories Arrive." *Seminar announcement and abstract*, Announcement available at https://sle.cueb.edu.cn/ljxs/bb3e37e8a2f44b9aa0344b8352e2caad.htm; accessed August 30, 2026.

Song, Yang, and Wei Zheng. 2016. "The Impact of the Sent-Down Movement on Chinese Women's Age at First Marriage." *Demographic Research*, 34(28): 797–826.

Tian, Felicia Feng. 2016. "Transition to Adulthood in China in 1982–2005: A Structural View." *Demographic Research*, 34(16): 451–466.

Xie, Yu, and Jingwei Hu. 2014. "An Introduction to the China Family Panel Studies (CFPS)." *Chinese Sociological Review*, 47(1): 3–29.

Xiong, Wanru. 2023. "Love is elsewhere: Internal migration and marriage prospects in China." *European Journal of Population*, 39(1): 6.

A Crosswalk Construction and Validation

The census files identify geographic units with survey-specific numeric codes and English value labels; the Fan–Zou data identify 1982 prefectures with four-digit codes and names. We normalize labels by lowercasing, removing administrative suffixes, and collapsing whitespace, and permit a match only when the first two digits of the prefecture code equal the census province code; combined geographic units in the IPUMS labels are excluded, and each census code must map to at most one prefecture. Name-only matching would admit 13 false candidate matches in 1982 and 10 in 1990—same-named units in different provinces—which the province constraint removes. The constrained match leaves 188 matched 1982 prefectures (157 with a valid heavy share) and 233 in 1990 (189 valid); intersecting the 1982–1990–2000 supports leaves 154, and dropping single-prefecture provinces leaves 148 in 21 provinces. The validation suite enforces crosswalk uniqueness, fixed support, sample counts, artifact freshness, and the reported numerical identities; implementation details appear in `REPLICATION.md`.

B Additional Census Results

B.1 Cross-sectional age-bin slopes and grouped estimates

The DDD of Equation (2) is built from three cross-sectional slopes. In the common 1982 geography, the Third Front differential in the 20–24 heavy-share slope is 4.21 points per standard deviation when ages are estimated separately, against 1.06 at 25–29 and approximately zero at 30–34; removing students and collective dwellings gives 4.38, 0.89, and zero; education adjustment gives 3.09, 1.38, and 0.34. In 1990 the young cross-sectional differential is 2.10 points and in 2000 it is 3.76; the same-major-unit 2000 restriction gives 2.99. These estimates do not absorb prefecture levels, which is why the body emphasizes Equation (2). Figure 5 displays the grouped DDD estimates on the identical geography across census years.

Table 6: Functional-form and province-designation diagnostics

Specification or contrast	Estimate	Standard error	Diagnostic p
Panel A. Exact reassignment of Third Front labels among interior provinces			
Baseline age-profile DDD	4.23	1.66	0.059
Predetermined siting controls	2.24	1.15	0.249
Panel B. Nonlinear exposure diagnostics			
Below-median spline slope	2.52	1.35	0.078
Above-median spline slope	2.88	2.73	0.305
Above-minus-below slope	0.36	3.54	0.920
Joint differential-linearity test	$F_{2,20} = 0.41$	—	0.667
Top-minus-bottom exposure quartile	8.99	4.07	0.039
Joint young-age quartile test	$F_{3,20} = 1.73$	—	0.193

Notes. Panel A reports percentage points per common-prefecture standard deviation and conventional province-clustered standard errors. Its p-values enumerate all 715 ways to designate nine of the 13 interior provinces while holding coastal provinces undesignated. Historical designation was not uniform over these assignments, so this is an exact conditional sensitivity diagnostic, not causal randomization inference. Panel B uses the baseline design. Spline slopes are percentage points per 0.10 increase in the heavy-industry share; the knot is the common-sample median. Quartile entries compare the indicated Third Front age-profile contrast with the bottom exposure quartile. The top-quartile coefficient is largest, but the bin pattern is not monotone and its joint test is imprecise.

B.2 Composition outcomes and sex-specific estimates

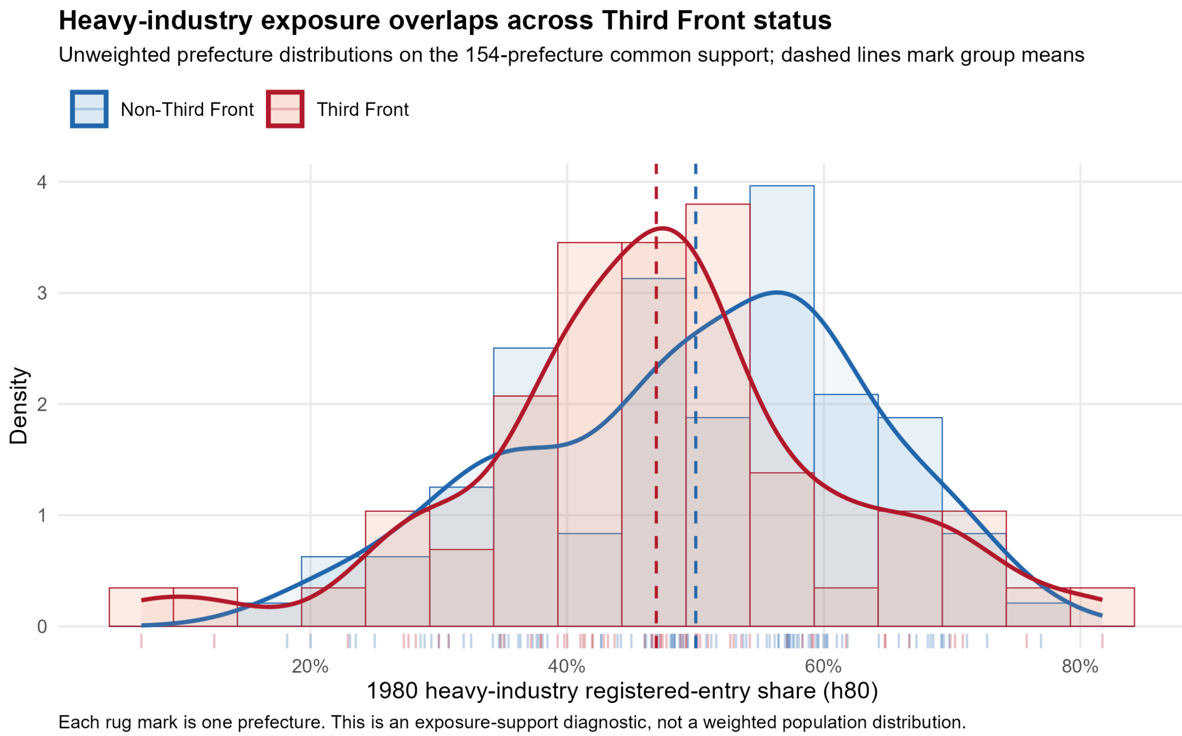
Table 7: Composition outcomes and sex-specific marriage estimates

	Ages 20–24		Ages 25–29	
Panel A. Composition outcomes, same DDD (1982)				
Currently a student	+0.24	(0.23)	+0.01	(0.04)
Collective dwelling	−0.37	(0.36)	−0.27	(0.21)
Secondary schooling or more	+4.33***	(1.43)	+1.57**	(0.62)
Panel B. Composition outcomes (2000)				
Currently a student	+2.11*	(1.14)	+0.18***	(0.06)
Collective dwelling	+3.31*	(1.88)	+0.52	(0.49)
Secondary schooling or more	+0.59	(1.15)	−0.16	(0.43)
Moved province within five years	+1.65	(1.25)	+0.14	(0.44)
Move reported as marriage-related	−0.42	(0.26)	+0.28	(0.30)
Panel C. Never married, by sex (1982)				
Women	+4.67**	(2.17)	+0.62	(0.54)
Men	+3.79**	(1.56)	+1.55	(0.95)

Notes. Each row applies the core design of Equation (2)—the same support, age contrasts, fixed effects, weights, and province clustering—to the indicated outcome (Panels A–B) or estimates the marriage outcome within sex using province-by-age effects (Panel C). Entries are percentage points per cross-prefecture standard deviation of the 1980 heavy share, contrasted against ages 30–34. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

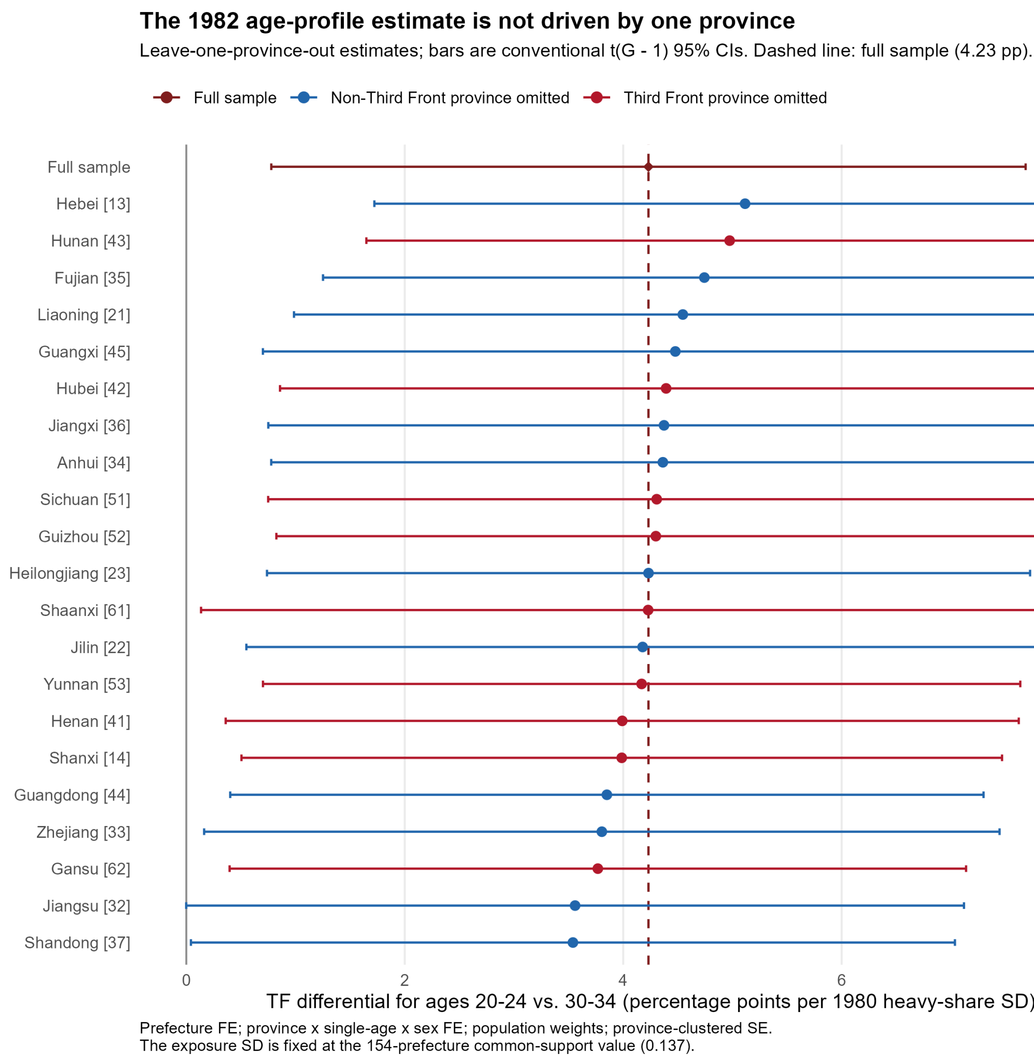
B.3 Exposure support and province influence

Figure 3: Overlap in the 1980 registered heavy-industry share



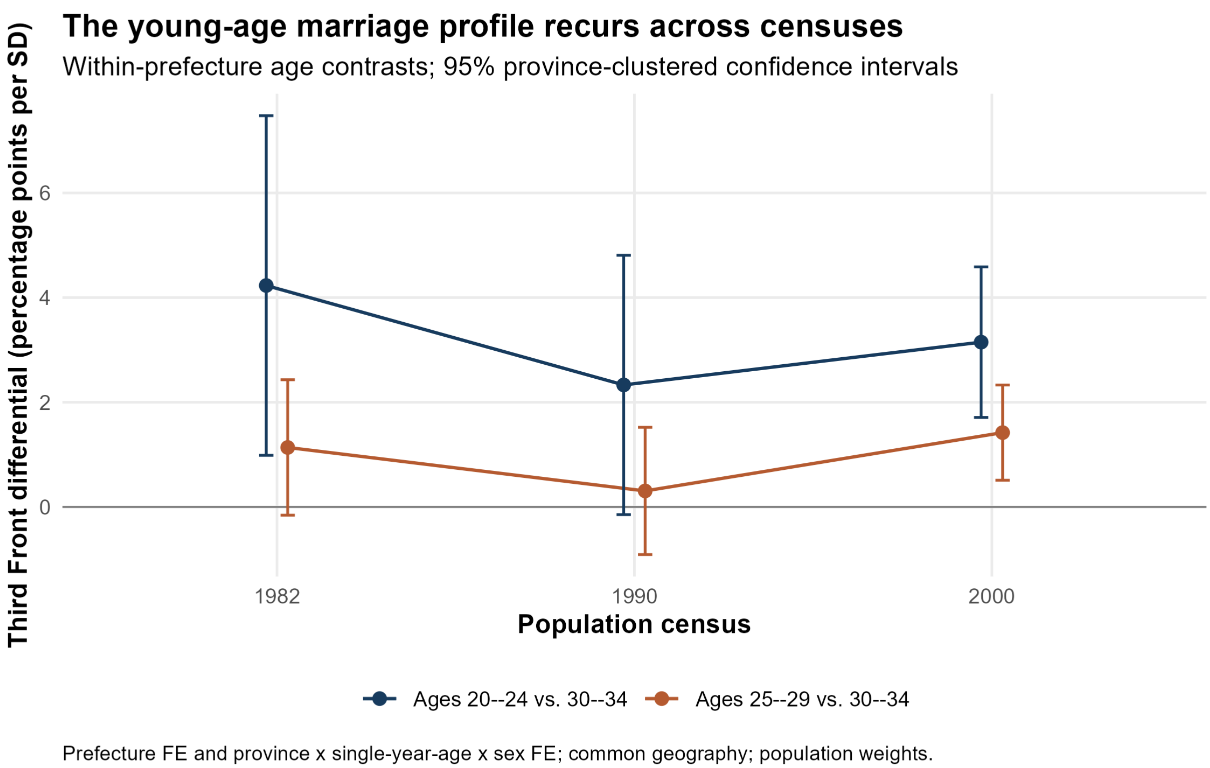
Notes. Unweighted prefecture distributions on the 154-prefecture common geography. Each rug mark is one prefecture; dashed lines are group means. The share is computed from cumulative registered entries, not contemporaneous plant employment.

Figure 4: Leave-one-province-out influence diagnostic



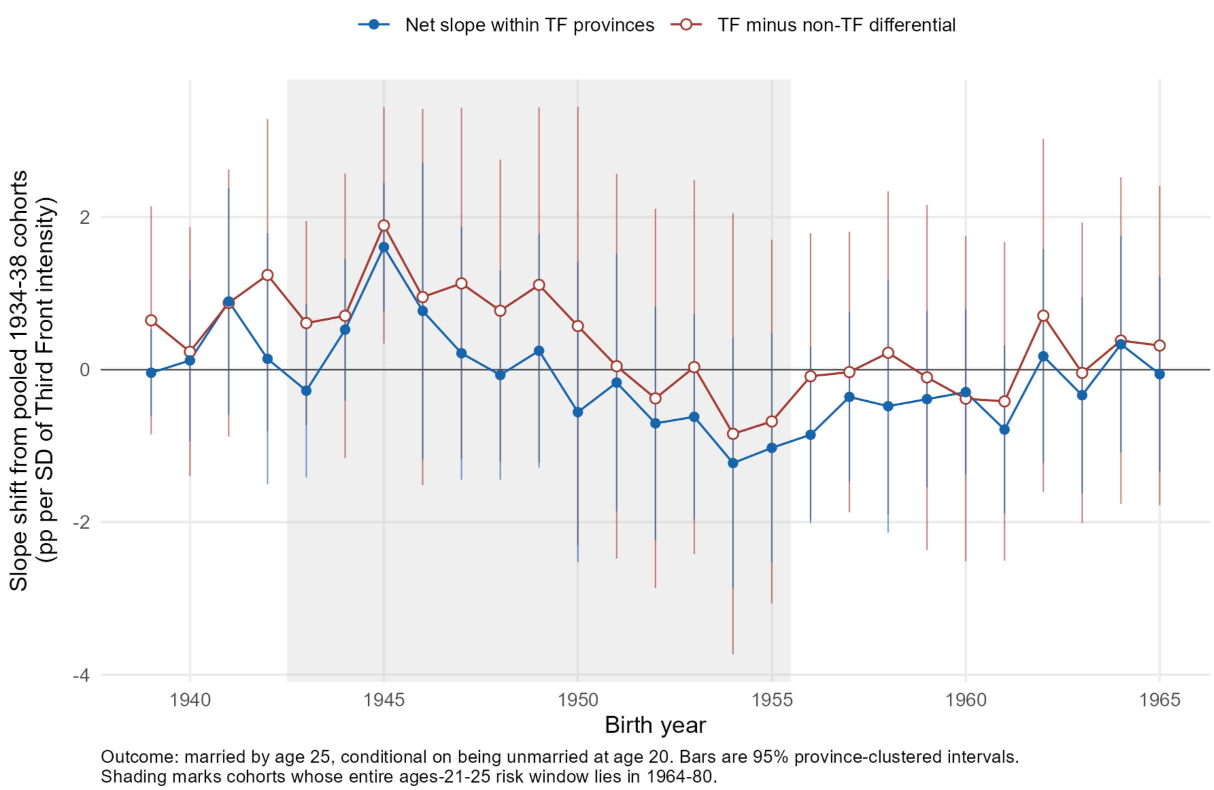
Notes. Each row re-estimates the 1982 baseline after removing the named province; the full-sample result appears at top and as a dashed vertical line. Intervals use conventional province-clustered standard errors and a $t(G-1)$ critical value. The exposure scale remains the standard deviation on all 154 common prefectures. This is an influence diagnostic, not a cure for endogenous siting.

Figure 5: The age-profile differential across censuses



Notes. Estimates reproduce the baseline rows of Table 2. Error bars are conventional 95-percent confidence intervals based on province-clustered standard errors. All three censuses use the same matched prefectures and fixed-effect structure.

Figure 6: Birth-year shifts in marriage by age 25



Notes. Coefficients are changes in the heavy-industry slope relative to pooled 1934–1938 birth cohorts for marriage by age 25 among people unmarried at 20. The figure reports the slope outside designated provinces, the additional Third-Front differential, and the net slope inside designated provinces. Shading classifies each cohort's ages-21–25 risk window relative to the 1964–1980 build-out. Models include prefecture fixed effects and province-by-single-birth-year-by-sex effects, use population weights, and cluster by province. Exposure is assigned by prefecture of residence in 2000; the figure is a timing diagnostic, not an event study with pre-exposure residence.

B.4 Direct project measures and the railway IV

Table 8: Direct Third Front exposure and the 1982 age profile

Treatment	Specification	20–24 vs. 30–34	25–29 vs. 30–34
Third Front investment	Unadjusted	1.61 (1.07)	−0.19 (0.35)
Third Front investment	Siting adjusted	0.46 (1.13)	−0.02 (0.35)
1985 plant employment	Unadjusted	2.34 (1.17)	−0.16 (0.32)
1985 plant employment	Siting adjusted	0.92 (1.37)	0.12 (0.43)
Designated industrial center	Unadjusted	1.20 (2.29)	0.94 (0.58)
Designated industrial center	Siting adjusted	3.93 (2.36)	2.01 (0.76)

Notes. Percentage-point age contrasts per treatment standard deviation, except that the center indicator is a zero-to-one change; conventional province-clustered standard errors are in parentheses. Estimates use 34 Fan–Zou hinterland prefectures in only seven effective provinces, prefecture fixed effects, province-by-single-age-by-sex effects, and population weights. With seven clusters, conventional tail probabilities are unreliable; the table therefore reports no significance stars and is descriptive. Siting-adjusted rows interact 1980 rail distance, mining capacity, terrain, elevation, and distance to the provincial capital with age bins. The direct-investment first stage on 1962 rail distance is weak ($F = 2.39$ under core controls); the plant-capacity first stage is $F = 9.58$.

Table 9: Plant capacity and the 1982 age profile: railway-IV diagnostic

Age contrast	Specification	OLS	2SLS	First-stage F	Wild AR p
20–24 vs. 30–34	Baseline	1.64 (0.98)	1.98 (1.52)	5.1	0.340
25–29 vs. 30–34	Baseline	−0.01 (0.31)	−0.16 (0.37)	5.2	0.627
20–24 vs. 30–34	Siting controls	1.84** (0.84)	1.96** (0.87)	34.8	0.366
25–29 vs. 30–34	Siting controls	−0.00 (0.24)	−0.18 (0.35)	33.6	0.675

Notes. Coefficients are percentage-point effects of a one-unit increase in the Fan–Zou 1985 non-mining large/medium-plant employment share. The 1962 railway-distance interaction instruments the capacity interaction. All models contain prefecture and province-by-single-age-by-sex fixed effects, weight by population, and cluster by prefecture. Siting controls are 1980 rail distance, mining capacity, slope, elevation, and distance to the provincial capital, each interacted with the age contrast. The overlap contains 22 prefectures. Wild AR p is a 4,999-draw restricted-null wild-cluster test of the reduced form and is robust to weak instruments.

Given the 22-prefecture overlap, high leverage, and weak-instrument concerns, the wild Anderson–Rubin test is the primary inference; the conventional 2SLS p-value is not.

B.5 Work-unit housing in the 1988 CHIP urban sample

Table 10: Industrial employment and public housing, CHIP 1988 urban sample

	Other household head	Industrial household head	Industrial gap
Non-Third-Front provinces	0.852	0.784	−0.069
Third Front provinces	0.870	0.902	+0.032
Raw difference-in-differences			0.101
Province-FE specification			0.110**
			(0.036)

Notes. Public housing pools employer-owned and municipal housing-bureau dwellings. The sample contains 8,094 households in ten provinces. The province-fixed-effect model controls for the head's sex and age and household size and clusters by province. Restricted-null wild-bootstrap $p = 0.0165$; exact designation-reallocation diagnostic $p = 0.0119$. The table establishes differential public-housing dependence, not the effect of a housing queue on marriage.

C CFPS Construction and Results

The seven CFPS waves from 2010 through 2022 contain 168,307 person-wave records in the access-controlled project extract (Institute of Social Science Survey, Peking University, 2015), collapsed to one row per person using modal stable characteristics and the earliest valid first-marriage year, birth cohorts 1930–1995, and valid survey city. Controls are sex, ethnicity, and parental education. Respondent education, used only descriptively, is the maximum valid completed years reported across waves. For a threshold a, a person enters if observed to at least a and unmarried at twenty; never-married people carry outcome zero, and people married before twenty are outside the risk set rather than assigned post-event exposure. The fixed-age risk sets contain 16,706 observations at 25 and 16,135 at 28. The regression interacts the city heavy share in the year the person turned twenty with designation, under city and cohort fixed effects.

Table 11 reports these fixed-age associations.

Table 11: CFPS fixed-age marriage associations

Exposure construction	Married by 25	Married by 28	Observations (25/28)
Cumulative-entry registry	−1.70** (0.81)	−2.00*** (0.65)	16,706/16,135
Surviving-stock registry	−1.40* (0.79)	−1.50* (0.74)	15,694/15,123

Notes. Entries are percentage-point changes per within-city standard deviation in the heavy-industry share at age twenty, for the Third Front slope relative to the slope elsewhere; city-clustered standard errors are in parentheses. Samples retain respondents unmarried at twenty, including those never married, and require that the threshold age was observed. Models include city and birth-cohort fixed effects plus sex, ethnicity, and parental education. Location is the survey-era city, so these are secondary associations rather than a historical treatment design.

A marriage-year specification—age at first marriage on industrial intensity in the marriage year—is retained only as a diagnostic: it conditions on having married, dates exposure with a year the outcome generates, and gives inconsistent estimates across the two registry constructions on the same respondents; moreover, survey age, marriage year, and age at first marriage satisfy an arithmetic identity within wave, so including survey age with marriage-year effects can mechanically explain the dependent variable. Its canonical interaction is 0.891 years (s.e. 0.455) under city clustering, with province wild-bootstrap $p = 0.096$ and exact interior-designation $p = 0.150$, and falls to 0.401 without city trends.

Matching CFPS cities to the Fan–Zou plant data yields 116 prefectures (21 hinterland). Capacity is associated with 2.70 points less marriage by 25 and 1.54 by 28 across all matched prefectures (wild prefecture-bootstrap $p = 0.060$ and 0.027); hinterland estimates are -2.98 and -5.95 with wild $p = 0.559$ and 0.216. Only three designated centers fall in the usable overlap; their indicator is associated with about five points less marriage at both thresholds, with wild $p = 0.082$ and 0.114. These are descriptive, few-site associations. We report no linked CFPS city names or below-province outputs. The railway IV in this overlap has first-stage F near five and uninformative 2SLS estimates.

D Reproducibility and Specification Status

The analysis was exploratory and not preregistered. We therefore distinguish the final estimands and inference procedures from specification history, which remains documented in the replication materials. The reproducible pipeline rebuilds the province-safe census crosswalk from value labels, constructs every analysis sample from raw or licensed source files, generates the manuscript exhibits, and writes machine-readable manifests for sample counts, exposure scales, bootstrap results, and decomposition identities. An automated fail-hard entry point stops on any unsuccessful stage; a separate validator checks artifact freshness, support, numerical identities, and manuscript claims. The complete script-to-exhibit map and data-access restrictions are recorded in `REPLICATION.md`.