Democratizing Opportunity: The Effects of the U.S. High School Movement*

Alison Doxey[†], Ezra Karger[‡], and Peter Nencka[§]

March 20, 2025

Click here for the most updated version.

Abstract

The U.S. high school movement transformed economic opportunities for young adults. We present a new complete panel of high schools in over 25,000 towns and cities across the United States. We use this data to show that high school access caused sharp increases in school attendance rates for high school-aged students but not younger children. Linking children to their adult labor market outcomes, we show that high school access boosted labor supply, increased job quality, and reduced the probability of marriage by age 28, with effects concentrated among women, who outnumbered men as high school graduates until the 1940s. We then use structured biographies to show that high school availability increased the probability that a child would become eminent, democratizing access to the most prestigious positions in science, politics, and business.

*Thanks to seminar participants for their helpful feedback at the Cliometrics Association, Economic History Association, NBER Development of the American Economy Summer Institute, Georgetown University, Mountain West Economic History Conference, Northwestern University, Marquette University, Ohio State University, Ohio University, University of Chicago, and the University of Wisconsin – Milwaukee.

[†]Harris School of Public Policy, University of Chicago

[‡]Federal Reserve Bank of Chicago

[§]Department of Economics, Miami University

1 Introduction

The late 1800s to 1940 marked a significant transition era in the United States. The country experienced a shift away from an agriculture-focused economy as industrialization and urbanization surged. This shift coincided with increased participation of women in the workforce. Between 1890 and 1940, the percentage of women aged 25 to 44 in the labor force grew from approximately 18% to 35%, reflecting a substantial change in gender roles (Goldin, 2006). The increase for older, married women was substantial: in 1890, almost no white married women aged 35-44 worked. By 1930, nearly 20% did (Goldin, 2006). This growth was fueled by the expanding demand for labor in offices and service industries, which created new employment opportunities and drove urban migration.

Rapidly changing technology and opportunities coincided with a "race" to educate the population at an unprecedented rate (Goldin and Katz, 2008). In this paper we study a critical institution that led to the fastest increase in schooling attainment in American history: the high school. In 1910, only 9 percent of 18-year-olds held high school degrees. By 1940, 50 percent did (Goldin and Katz, 2008). This rapid increase was fueled by a surge in local high schools, from fewer than 3,000 in the late 1800s to more than 20,000 by the early 1950s.[1] Accessible, publicly-funded, and tuition-free high schools catapulted the United States education system ahead of comparable European countries. Increased high school attendance disproportionately affected women, who graduated at higher rates than men until the 1940s (Goldin, 2008; 2021).

Despite recognition that high schools dramatically increased enrollment, it is difficult to estimate how high school access affected completed educational attainment and labor market outcomes. It is even harder to link the expansion of high schools to the transformative societal trends that defined America. Data is limited; to date, researchers have analyzed the expansion of public high schools across the U.S. using state-level summary reports of high schools or county-level

[1] The population of the United States also grew, but at a much lower rate. The number of school-age children in the United States increased by only 65% from 1890 to 1950, relative to a 560% increase in high schools.

statistics for select states (Goldin, 1998, 1999). However, access to schooling varied dramatically *within* states and counties over time, so it is challenging to map aggregate data to the exact opportunities available to residents. The potential effects of schooling are also challenging to measure. The 1940 census was the first to ask respondents to self-report educational attainment, and this outcome is measured with significant error since respondents vastly overstated their level of schooling. Roughly twice as many adults reported having attended high school as is possible, given administrative records describing high school enrollment rates for those cohorts (Goldin, 1998). Tracking long-run outcomes for women is particularly difficult. Because of name changes after marriage, until recently, it has been impossible to use linked census data to simultaneously observe their childhood location, careers, and demographic outcomes.

In this paper, we provide the first estimates of the effect of high school access on the long-run outcomes of both men and women. We construct a new, complete panel of all public and private high schools in the United States through 1951. The panel contains information about each high school's founding date and location, as well as counts of enrollment and teachers by school for a subset of years. We combine this panel with complete-count census data from 1850 to 1940, linking people across census waves using the Census Tree (Buckles et al., 2023). This process allows us to observe a large sample of both men and women in childhood and adulthood.

We estimate the effect of high school access on enrollment and adult outcomes, exploiting variation within similar towns and cities that expanded high school access at different times. Our approach builds on an extensive literature that links the construction of schools and school spending to short- and long-run local outcomes (e.g., Duflo, 2001; Duflo, 2004; Aaronson and Mazumder, 2011; Schmick and Shertzer, 2020; Navarro-Sola, 2021; Andrews, 2023). We estimate event studies that trace out the effects of high-school access by comparing students who were born slightly too early to be able to attend schools to those who had partial and complete access to high schools, given their age.

We find that the introduction of high schools led to an increase in the rate of youth attending

school and that this increase grew over time, consistent with aggregate enrollment trends. We see little increase in enrollment for children aged 13 and 14, consistent with the idea that many of these students would have been attending middle school or part-time schooling and would, therefore, have been relatively unaffected by high school entry. For 15–18-year-olds, we see positive enrollment effects that start in the years immediately after high schools open. For older students plausibly most affected by high schools, enrollment effects are largest: for example, after two decades, we see an increase in enrollment of about 20 percent for 17-year-olds.

Linking 16–18-year-olds to their young adult outcomes as 26–28-year-olds, we find evidence that high school entry boosted labor supply, but only for women. We also find that after high schools opened, men and women were more likely to enter higher-skilled occupations, such as clerical and professional positions. These shifts translate to increases in occupational income, with much larger effects for women. Consistent with high schools prompting a shift in career choices that required corresponding family choices, we observe reductions in the probability of women marrying by age 26–28. When we link forward another ten years and look at outcomes at ages 36 to 38, we see that most labor market effects persist in sign but attenuate in magnitude, particularly for women. This pattern of results suggests that while high schools expanded opportunities for young women, many later left the labor market to marry.

To explore our labor market outcomes in more detail, we use a dataset of notable men and women drawn from crowd-sourced structured biographical sketches from Wikidata. These biographical sources include exact dates and birthplaces for both men and women. Our outcomes include various indicators for notability, including whether a child grew up to be a federal judge, congressperson, eminent scientist, recognized businessperson, or artist. These are right-tail human capital outcomes, and unlike in most historical datasets, we can measure them consistently throughout our entire sample period for both men and women. Linking high schools to our panel of notable people, we find that access to a high school increased the probability that a child became eminent as an adult, with larger percentage increases for women.

Related Literature

This paper makes several contributions. First, we build on the work of economists and historians describing the high school movement and contemporaneous increases in school spending and investment. Goldin (1998) documents the dramatic expansion of high school enrollment and graduation rates between 1910 and 1940, particularly in non-Southern states. Goldin provides one of the first examinations of the correlates of early expansion, showing that investment in high schools was associated at the state level with greater levels of wealth and lower levels of manufacturing. Goldin (1999) expands on this work with additional data linking the high school movement to the education wage premium, showing that as high schools expanded across the U.S., the high school wage premium fell by 37 log points from the 1890s to 1939. Goldin and Katz (2008) further explores the interplay between the secondary school movement and the U.S. wage structure. Schmick (2024) and Schmick and Shertzer (2020) study the determinants and effects of school spending increases in the post-WW1 era in urban cities, and Cascio and Lewis (2024) show that increasing Black teacher salaries in the South in the 1930s had positive effects on students. We focus on capital investment in high schools and link high school access to both the short- and long-run outcomes of affected children. We also concentrate on women, for whom it has only recently become feasible to study linked child-adult records.

Second, our work relates more broadly to modern and historical work that links the construction of schools and school spending to long-term outcomes. Duflo (2001) pioneered this empirical design, using the construction of schools across Indonesia in the 1970s to measure the effects of educational access on wages. A sizeable recent literature, summarized in Jackson and Mackevicius (2024), measures the effects of school capital investment on short-run (test score) and long-run (educational attainment) outcomes. Our work provides a historical complement to this work, focusing on one of the formative expansions of educational opportunity during a period of industrial transformation.

Third, we estimate the effect of high schools on a broad range of occupations. By linking to

datasets of eminent people, we estimate whether high schools shifted the probability of observing children who grow up to have occupations in the right tail of the human capital distribution. We contribute to an emerging literature examining the predictors of growing up to be an innovator, scientist, and related occupations, outcomes that play a crucial role in driving innovation, economic growth, and inequality (e.g., Bell et al. 2019; Hsieh et al., 2019; Aghion et al., 2019).

2 Data

2.1 High Schools

We construct our panel of high schools in the U.S. from the 1800s to the mid-1900s by combining information from four sources. First, we compile information from censuses of all public and private high schools, collected by the Bureau of Education every 1–2 years from 1873 to 1905 and in 1912. After 1912, the next census of high schools was published in 1951. These censuses contain information about the number of students, the number of teachers, the length of study, and sometimes the founding year of each high school.[2] We use founding years to retrospectively construct a consistent panel of high schools across the United States, even for years when high schools are not surveyed.[3]

For later years, we use lists of accredited high schools published by the Bureau of Education every 2–6 years from 1911 through 1944. Accreditation standards varied by state, and public universities used published bulletins to offer admission to local students with a diploma from a high school that met a set of criteria. These criteria often included a requirement (1) that the high school offers four years of study, and (2) that the high school offers at least a minimum number of math, English, and history credits. These lists of accredited high schools contain no information

[2]The founding year was only collected in a few of the censuses, including in 1903.

[3]A limitation of this approach is that we will not observe high schools that closed using retrospective data. In our empirical analysis, we do not focus on individual high schools but on the existence of any school within a city. A city rarely loses access to a high school once built, so we do not believe this is a limitation for our analytical purposes.

on founding dates or enrollment, but they confirm the existence of public and private secondary schools in particular cities and towns.

To complement the founding dates from the earlier high school censuses, we use *Patterson's American Educational Directories*. These provide names of high schools in each county for seven years between 1906 and 1924. The Patterson's directories contain establishment dates for reporting high schools. However, these directories were less comprehensive than the contemporaneous censuses of high schools produced by the Bureau of Education [4]

To supplement our data on private high schools, we collect information from the more recent Private School Universe Surveys produced by the National Center for Education Statistics. These surveys include the founding dates of reporting schools. We augment our historical panel with the founding dates of high schools reported in the 1989 or 1995 academic year surveys.

We digitize each source except the more recent Private School Survey, which is already machine-readable. We combine these four sources into one panel of high schools containing 360,000 school-by-year observations. The median observation in our panel comes from 1922, but the critical data for our analysis come from the 1889 through 1905 high school censuses, which are available annually.

To prepare this panel for analysis, we:

1. Identify transcription errors in the names of towns and cities and correct those errors by looking for places that exist uniquely in the panel, are one letter away from other place names that appear in the panel and do not conflict on available years.

2. Collapse data to place-year level, calculating enrollment of local schools and the earliest reported founding year of public and private schools.

3. Link each high school's town or city name to GNIS place data, Geocodio, Texas A&M's geocoding platform, and the HERE Geocoder API to identify a longitude-latitude corre-

[4] We use Patterson's directories for the 1906, 1908, 1912, 1913, 1914, 1920, and 1924 academic years.

sponding to the location of each school's town or city. This step involves standardizing place names and relying on county-level and state-level location identifiers in the high school data when available. We use multiple geocoding platforms to look for agreement and disagreement across sources.[5]

4. We take all of the potential geocodes for each place and address discrepancies by privileging geocodes that appear to be within five miles of each other independently in multiple sources. We also require that the implied geocode is within the United States.

At the end of this process, we have a panel of 26,877 places with information about the longitude and latitude of the place where we observe that high school. Figure 1 maps the spread of high schools over time. Towns and cities with high schools by the 1800s are concentrated in New England and the Midwest. High schools expanded later in the South, and towns and cities in the West also lagged New England in schooling investment.

Figure A1 shows that despite these general patterns documented in past work, aggregate state-level high school access measures hide substantial heterogeneity within states. For example, comparing Panels A and D, we see that the average city in Massachusetts received a high school much earlier than the average city in Tennessee. But *within* each state, there is substantial variation in the timing of high school access. This within-state and county variation is particularly evident in the Midwest, as shown in Panels B and C for Ohio and Wisconsin.

In Panel A of Figure 2, we plot the increase in the number of towns and cities in the United States with at least one high school between 1800 and 1950. We show the cumulative number of cities with a public high school and 'any' high school, which includes private high schools.[6] While

[5] In this step, we can encounter difficulties with non-unique matches if two cities or towns with the same name are in a given county or state. We flag these observations for exclusion in robustness checks.

[6] Note that we plot the earliest year when a high school could have existed in our town, according to our panel. So for example, because we have lower-quality data between 1905 and 1910, we assume that each town that reported a high school in 1910 had access to that high school as of 1906–the first year when our panel implies that a high school could have existed in those towns. This approach creates discontinuities in the growth rate in Figure 2. Our results are robust to excluding high schools assigned to these discontinuous years.

some towns and cities built high schools in the early- to mid-1800s, we see the largest increase in high schools between 1880 and 1910, when the number of towns with high schools more than tripled. After 1910, we no longer have access to higher-quality high school census data. The flat growth rate of the number of towns with a high school between 1910 and 1915 is driven by a lack of data. The subsequent upswing in the number of towns with a high school in our panel is caused by the large number of accredited high schools reporting to the Bureau of Education for the first time in the 1920s and 1930s. Because of these changes in data quality, in our main results, we focus on variation in high school construction before 1909 and drop cities that established high schools after 1909 from our analysis.[7]

In Panel B of Figure 2, we plot the total enrollment of students in high schools in years of our data with valid enrollment information. Consistent with prior work by Goldin (1999), who reports Bureau of Education aggregate graduation and enrollment information from 1890 to 1970, we find a relatively linear increase in enrollment rates from 1890 to 1910.[8] In years where we have enrollment information aggregated from the high school panel and Goldin (1999) reports aggregate enrollment rates, our reported enrollment rates are similar. For example, in 1910, Goldin (1999) shows a high school enrollment rate of 14–17-year-olds of around 15% , and there were roughly 7 million 14–17-year-olds alive in 1910, implying an enrollment count of 1.05 million students. Aggregating our high school panel, we estimate that 1.06 million students attended high schools in 1910.[9]

[7]The analysis sample includes any high school before 1909 and those that never established a high school before 1950. We include cities that established a high school before 1850 in our analysis sample, and they can contribute to the estimation of post-period coefficients. However, since 1850 is our first census year, they do not contribute to pre-period estimates. Our results are robust to excluding pre-1850 cities from our sample.

[8]Goldin (1999) finds a roughly linear increase in graduation rates from 1890 through 1900 and a linear increase in attendance rates from 1900 through 1910 when the Bureau of Education begins to report this information.

[9]Note that in Figure 2, we report this total in 1912, which was the year of the Bureau of Education Report, but which refers to the 1910 school year

2.2 Census

We use the historical decennial U.S. censuses to track enrollment and adult outcomes (Ruggles et al., 2021). Each census after 1850 asks about school attendance, though the format and reporting patterns of those questions change over time. We append the 1850-1930 censuses together and focus on enrollment for 13–18-year-olds in the year of each census. Each adult reports their occupation, industry, and marital and household status. We link high schools to the census using the Census Place Project, which converts census location strings into consistently geocoded cities, towns, and unincorporated places (Berkes et al., 2023).[10]

Our baseline analysis assigns a city as treated once it has a high school within two miles of the Census Place Project geocode. To create a more balanced panel of cities, we use cities appearing in at least six possible eight censuses between 1850 and 1930.[11].

2.3 Census Links

We link individuals we observe in childhood to their adult outcomes in subsequent years. Unlike in some modern administrative datasets, linking is a non-trivial task because there are no unique identifiers across census years. Recent innovations in the historical record linking literature allow us to track both men and women over time and across places.

Our baseline analysis uses links from The Census Tree Project (Buckles et al., 2023). The Census Tree provides decade-to-decade links generated through several methods. The initial set of links comes from user-provided records on FamilySearch, a free genealogical website. Buckles et al. (2023) use these links as training data to generate additional links with a supervised machine learning algorithm. They supplement these links with connections from the Census Linking Project

[10]Throughout the rest of the paper, we will refer to these places as "cities," "towns," and sometimes simply as "places." For our purposes, all of these terms refer to this same type of sub-county location described above, which the Census Place Project assigns to an exact latitude and longitude.

[11]The 1890 census is not available since it was lost. Imposing a balanced panel removes cities founded during the greatest high school expansions in the late 1890s through the 1920s

(CLP; Abramitzky et al., 2022), the IPUMS Multigenerational Longitudinal Panel (MLP; Helgertz et al., 2023), and "hints" generated by algorithms on the FamilySearch website. Candidate links from these methods are adjudicated and provided as decade-to-decade crosswalks.

To construct our sample, we start with the universe of 16–18-year-olds in each census year since we are most likely to observe meaningful high school enrollment for this group. We use the Census Tree to link respondents forward to their outcomes when they are aged 26-28 and 36-38. We also link 16-18-year-olds to their age 6-8 observations to track childhood mobility. In robustness exercises, we consider wider age ranges.

Relative to other potential linking methods, Census Tree links are notable for their excellent coverage of women. Deterministic, name-based approaches like the CLP work well for men and can work well for linking older women. However, since women often change their names when married, these methods cannot be used to link girls forward to adult outcomes. The IPUMS Multigenerational Longitudinal Panel (Helgertz et al., 2023) includes high-quality links for a subset of women who stay in the same household as a linked male relative (e.g., a father or a husband). Similar to CLP-style methods, these links produce a selected sample when used to link women to their childhood homes. The genealogical data in the Family Tree dataset provides a unique and powerful way to link large numbers of teenage women forward, even as they change names, move households, and start careers.[12]

2.4 Notable People

Our data describing notable people come from Wikidata—a crowdsourced database of eminent people, including more than 100,000 U.S.-born eminent scientists, politicians, business people, and artists. This data describes the birthplaces and lifetime achievements of eminent men and women across the U.S.

[12] A limitation of many linking methods, including the CT, is that they include relatively fewer Black individuals. Therefore, we do not split results by race in this paper. We anticipate being able to estimate heterogeneity by race and region in future iterations of the project.

Wikidata is an attempt to produce accurate, structured knowledge about the world. It is crowd-sourced and maintained by a global network of volunteers. Most Wikidata entries are generated through corresponding edits on other Wikimedia projects, most importantly Wikipedia. Our analysis relies on a small number of characteristics (birthplace and sex) that are accurately recorded for eminent people. We clean the data from Wikidata and create a birth-year-by-city panel that includes information on the number of eminent people born each year in each place, by gender. We expand the panel and fill in zero-eminent-birth observations to create a balanced sample of birth years from 1810 forward. After geocoding the places in Wikidata, we link this panel to the location of high schools and calculate high school entry years for each city in the panel.

In Figure A2, we group our Wikidata panel by decade of birth and separately plot the number of men and women in our sample, showing that the number of eminent men and women we can track increased throughout the late 1800s and early 1900s. Because of their lower base rates in the early 1800s, women tracked in Wikidata grow at a faster percentage rate than men, though the absolute gap between men and women grows over time.

3 Census Results

This section describes our results from census data, proceeding in two steps. First, we estimate the effects of high school entry on contemporaneous school enrollment. This analysis is a "first stage" and tests whether local high school formation affects enrollment. Second, we use our linked sample to estimate analogous models for labor market and related outcomes for older adults whose childhood location we identify from the census links described above.

3.1 School Enrollment Outcomes

To estimate the effects of high school entry on enrollment, we specify an event study model of the form:

$$Enrollment_{ict} = \sum_{a=-30}^{60} \beta_a 1(HS)_{cab} + \delta_c + \eta_t + \lambda_b + \gamma_{s,t} + \varepsilon_{ict} \quad (1)$$

where $Enrollment_{ict}$ is an indicator for child i, city c, and decennial census year t who reports being enrolled in school. The indicator variable $1(HS)_{cab}$ tracks whether a child born in year b in place c, has access to a high school at age a; and δ_c, η_t, λ_b, and $\gamma_{s,t}$ are city, year, birth-year, and state-year fixed effects respectively. The coefficients of interest are the vector of coefficients β_a, which track enrollment effects as a function of the age when each high school opened in a town. We cluster standard errors at the city level throughout.

The intuition of Equation 1 is similar to that of a standard event study, but while the running variable is the relative age of each cohort when a high school entered, the outcome is the enrollment of that cohort when they were school-aged. To illustrate, start by focusing on the outcome of enrollment at age 16–18. If you were 25 when a high school opened in your childhood town, it would be impossible for that high school to affect your enrollment when you were 16–18. When we plot coefficients, if we see an increase in the predicted probability of enrollment if a high school opened when you were 25 compared to when you were 35, that is a pre-trend: it suggests that some other characteristic of the town was changing across those cohorts that affected children's enrollment rates. If a high school opens during your later teenage years, you could be partially treated by the school. If the school opens before you are 14, you will be fully treated because you can attend the high school for the full range of traditional schooling years. This treatment and partial treatment definition can be fuzzy. Some older children might attend high schools (especially when they were first built and may have had fewer age restrictions), so older cohorts can still be partially treated. In the following baseline analysis, we estimate (1) using three-year age bins, and the omitted category is youth who were 21–23 when schools opened.

As is now well known in the literature, this type of event study estimation can be biased if there are heterogeneous treatment effects across treated cohorts over time. Since event studies

like Equation 1 use early-treated units as controls for later-treated units, treatment effect dynamics can introduce bias.[13] To address this concern, we also estimate a version of Equation 1 that uses the two-stage difference-in-difference imputation procedure outlined in Gardner (2021). In the first stage of this process, we identify city, year, birth-year, and state-year fixed effect coefficients using a sample of untreated and not-yet-treated observations. We then residualize our enrollment measure ($Enrollment_{ict}$), subtracting off the predicted enrollment from the first stage equation, and estimate Equation 1 using this residualized variable as our outcome in a simple event-study regression without covariates.[14]

Figure 3 shows the results of this estimation for a pooled sample of respondents who report being 16–18-year-olds in each census year for the standard, two-way fixed effect estimator (Panel A) and the two-stage Gardner approach (Panel B). We focus on this age because these students were plausibly the most likely to continue their schooling once a high school was built. Effect sizes grow rapidly in absolute and percentage terms for high school-aged students with access to schools. For example, effect sizes grow using the Gardner method to over a 5 percentage point increase in the probability of attending any school, or roughly a 15 percent increase relative to 1870 enrollment rates. This percentage change likely understates the actual percent increase in high school enrollment. Thirty-two percent of youth aged 16–18 reported any schooling attendance in the previous year, but we know that very few of them were graduating high school – less than 5 percent did by 1880, according to aggregate records (NCES, 2010). This discrepancy is due to the question on the census: enumerators were asked to report if a child attended school during the year, not high school. Many may have attended junior high schools or smaller schooling facilities part-time, but we would not consider them as attending high schools.[15] Using the aggregate grad-

[13] For a discussion of this bias and various remedies, see Baker, Larcker, and Wang (2022).

[14] Relative to other proposed methods to overcome issues in staggered difference-in-difference settings, the Gardner (2021) approach is extremely flexible and extends easily to our event study setting with covariates.

[15] The 1870 enumeration instructions do not distinguish between types of school, except to note that the census question "is not intended to include those whose education has been limited to Sunday or evening schools." It is unclear how individual enumerators asked the question.

uation figure as baseline probability would imply a much larger first-stage increase in schooling but would overstate the impact since some students enticed to attend school will not finish. The actual change is likely between that number and using reported census enrollment to calculate the baseline probability.

The results in Figure 3 are similar using the Gardner and two-way fixed effect estimator; we use the Gardner method for the remainder of the paper. Figure 3 shows the corresponding estimates from this pooled sample of 16–18-year-olds, and Figure 4 shows estimates separately for men and women. Men's and women's enrollment responds similarly in the immediate decades after a school opens. However, on a percentage-point basis, the male enrollment effect is larger in the later years of our sample.

The estimates in Figure 3 and Figure 4 abstract from important heterogeneity by age. We show results separately for each age from 13 to 18 in Figure 5. For all ages, we see relatively flat pre-trends before high school entry, suggesting that towns that did and did not establish high schools were not on differential enrollment trends before entry. After high schools are founded, we see heterogeneous effects by age. We observe almost no increase in enrollment for 13–14-year-olds, and more delayed evidence of an effect on the school attendance rates of 15-year-olds. These results suggest that school enrollment for elementary and middle school children was less affected by high school entry, as expected.[16] By contrast, we see the largest, consistent increases in enrollment for 16, 17, and 18-year-olds. These are precisely the students for whom we would expect to see the largest effects: while junior high schools and middle schools could enroll 13-15-year-olds, it was less likely that older students would attend those schools. Given the differences by age in Figure 5, through the rest of the analysis, we show results for a consistent sample of 16–18-year-olds most plausibly affected by high school entry.

The growing effects that we observe in Figures 3, 4, and 5 imply that the availability of a

[16]High school entry could affect enrollment at earlier ages if it changes the perceived value of finishing earlier school. But we expect this impact to be smaller than the direct effect for older students

high school option did not increase enrollment dramatically in the immediate years after schools opened. This finding is consistent with past work on the high school movement, which showed the largest enrollment effects in later years. Because we focus on the construction of high schools before 1909, our median high school construction date is 1885. However, Goldin (1998) shows that the high school enrollment movement expanded most dramatically starting in 1910 before initially peaking around 1940. This finding is consistent with the timing of our effects. It suggests that places that initially established high schools were also places that later invested more in them and increased enrollment relative to places that did not have a high school nearby.

3.2 Adult Outcomes

To estimate the effects of high school access on adult outcomes, we estimate our Gardner specification with outcomes measured 10 and 20 years in the future for the same sample of youth we used in our enrollment results. Our linking strategy allows us to observe outcomes for people who stayed at and moved away from their childhood homes. Throughout, we cluster standard errors by the childhood city of people in our analysis sample.

Labor force participation

Figure 6 shows the results of our event study specification with an indicator for labor force participation as the outcome variable. The x-axis of these graphs shows ages relative to a high school opening for the 16–18-year-olds we link to these outcomes, so the corresponding results have the same interpretation as the first-stage results discussed in the previous section. For this outcome (and all following young adult outcomes), we show six different specifications: overall results at age 26-28 and age 36-38 and results separately by gender for both age groups.

Figure 6 shows modest increases in the probability of labor force participation in the overall sample. In the immediate aftermath of high schools opening we see an average increase in labor force participation of approximately two percentage points for 26-28 year-olds and 0.5 percentage points for 36–38-year-olds. But these aggregate effects hide important heterogeneity by gender:

panels (e) and (f) show precisely estimated null effects for men. By contrast, we see large and growing positive effects on labor force participation for women. After high schools open, women aged 26-28 are immediately more likely to be in the workforce. For women 9-11 when high schools opened, we estimate an approximate 2 percentage point increase in labor force participation. Relative to the 1880 mean participation rate for women in our linked sample, this is an 11 percent increase. Effect sizes for women grow over time, consistent with the first-stage effects on enrollment reported in Figure 4. Interestingly, the labor force effect attenuates for women by roughly half when looking at the 36-38-year-old sample. This decrease suggests that initial boosts to labor force participation are not fully maintained as women continue in their careers: we discuss this pattern in more detail as we examine other labor market and demographic outcomes.

Taking the female and male labor force results together suggests that high schools were an important contributor to the convergence in labor force participation observed in aggregate data during the first half of the 20th century.

Job choice and occupational income

To understand whether high schools affected the types of jobs that young adults entered in addition to their probability of any work, we estimate the same model with the probability of holding particular occupations as the outcome variable. We start with clerks, an occupation growing at this period and often requiring writing and arithmetic skills developed in high schools. The IPUMS clerk category includes bookkeepers, secretaries, and stenographers, which were among the most common skilled occupations that women began to enter in large numbers during this period.

In Figure 7, we show results with an indicator for reporting clerk as an occupation as our outcome variable. These figures show that the probability of being a clerk increased for affected men and women after high schools opened. The magnitude of the effect is larger for women than men—we see a 2 percentage point increase in the probability of reporting a clerk occupation for 26–28 year old women two decades out relative to an approximate 1 percentage point increase for men. This shift towards white-collar labor supports the argument in Goldin and Katz (2008) that

high schools imparted valuable skills (e.g., in mathematics and typing) to their students. Consistent with growing enrollment results, the effects grow for cohorts born in a city in the years after a high school opens.

Next, we test whether men and women were more likely to enter "professional" careers once they had access to high schools. This IPUMS occupation category includes doctors, professors, scientists, and engineers. Figure 8 shows increases in the probability of being employed in these professions for cohorts with access to a high school. Interestingly, we see similar percentage point effects for men and women for this outcome. However, the increase for women is a much larger percent change since only 1.3 percent of 36–38-year-old women in 1880 worked in a professional occupation, compared to 4.1 percent for men.

Similar estimates for teachers (Figure A3) show small increases in the probability of observing male teachers (off of a low baseline probability) and a larger increase in the probability of observing female teachers. However, Figure A3 shows that this effect on female teachers was strongest in the first two decades post-high school entry, after which it decreased and even became negative. This is a surprising pattern and could indicate that places that expanded high schools early were also places that later expanded other opportunities for women.

Figure 9 shows a summary of the main occupational results for men and women in the age 30s sample. This figure plots the coefficients from a pre-post difference-in-difference specification using the Gardner method. The coefficients capture the average effects of having access to high schools in the years after they open. Figure 9 shows that occupational reallocation from lower prestige and pay occupations like the service industry and manual labor towards clerical and white-collar industries occurred after high school entry for both men and women. However, the reallocation effect appears larger for men. In contrast, the labor force participation effect is larger for women, consistent with the results shown in the event study figures.[17]

To further summarize the combined effects of high schools on labor force participation and

[17]Figure A5 shows the equivalent figure for outcomes measured at age 26-28.

occupational choices, Figure 10 shows results with logged imputed occupational income as the outcome variable. Actual income is only available in the 1940 census. For this analysis, we use IPUMS's 1950 occupational scores, which assign median 1950 total income to similar occupations back in time.[18] Occupational income increases by roughly 10% for 26-28-year-olds in the cohorts born with high school access, though larger effects for women pull up this average. Interestingly, while the effect for men is a stable 4–5 percent increase in occupational income in their 20s and 30s, the effects for women attenuate by their 30s. This result is consistent with the similar attenuation observed in women labor force participation and occupational choice results.

Marriage

Our labor force participation and occupational choice results suggest that high school access may have led young adults, especially young women, to delay marriage. Indeed, our estimates in Figure 11 show evidence that this is the case; the share of married people aged 26–28 decreases. We see little short-run change for men; women drive the delay. By age 36–38, it appears that the delay is mostly over – while there is a slight decrease in the probability of marriage for women at that age, it is difficult to distinguish from a mild pre-trend in the years leading up to high-school entry. This finding is consistent with the attenuated occupational and income results for women in their 30s. While high school access appears to have led to a lasting boost in earnings for some women, many married and left the labor force in their 30s.

3.3 Robustness

Our analysis above shows that access to high school increases high-school-age enrollment and that students who had the opportunity to attend these schools were more likely to work and, conditional on working, more likely to work in higher-skill careers.

The primary identification concern with this analysis is that high schools do not solely drive

[18] We calculate logged occupational income as $ln(occupationalscore + 1)$ to handle non-working respondents, though other transformations produce similar results

the effects we observe. Instead, those effects might be caused by other differences between places that do and do not expand schooling opportunities for children. Time-invariant differences across cities that build and do not build schools cannot explain our findings: We do not see evidence of pre-trends. For example, someone who was 25 when a high school opened in their hometown is not likelier to become a clerk than someone who was 35 when a high school opened in their hometown. Instead, we only see increases in the probability of becoming a clerk for cohorts young enough to plausibly have high school access relative to their older peers who could not have attended these schools. This pattern suggests that the primary identification concern in our case is any city-specific, time-varying feature of a city that would affect economic development and opportunity for the exact cohorts affected by expanding access to high schools.

In Figures A6 through A13, we report a series of robustness checks on aspects of this potential concern for our enrollment outcome as well as for each of our career and demographic outcomes.[19] Panels A reproduces the baseline Gardner estimation research for each outcome. Panel B presents the corresponding two-way fixed effect estimate for each outcome — results are similar in this case across both estimators. Panel C shows two-way fixed effect results where we additionally condition on birth-year-by-state fixed effects, controlling for any birth-year-specific trends that might non-linearly correlate with both high school access and labor market outcomes. The results are similar to our baseline specification.

Panel D of each figure adds controls for year-by-county fixed effects to the two-way fixed effect model: this demanding specification controls for all time-varying, county-specific features that might correlate with high school entry and affect downstream outcomes. Results are similar to our baseline specifications across outcomes. Panel E shows results for a subset of 16–18-year-olds we can link to their childhood locations at ages 6-8. We condition on the set of respondents who have moved less than 15 miles between ages 6-8 and 16-18. This helps address a concern of

[19]For the career outcomes, we report combined results for both men and women over age 26-28 to minimize the number of total graphs. Other results are available upon request.

selective migration: our baseline effects could be driven by families who moved to an area *because* it had a high school. If that migration also correlates with other downstream outcomes, we could mistakenly attribute some of our results to it. We see noisy but similar patterns of results using the subset of non-moving households. Last, in Panel F we show a set of results where we broaden the sample to include 14-15 year olds instead of focusing only on 16-18 year olds. As expected, we see attenuated enrollment effects for this age group but similar long-run effects across outcomes.

4 Notable People Results

We next turn to the notable people results using Wikidata. The Wikidata panel is organized as a city-by-birth-year panel. This is a different structure from the analysis using the census data since we are not using individual-level data, instead aggregating to the birth-cohort-place level.[20]

After linking this panel to high schools, we calculate the age of each cohort when a high school first opened in their city. We use this age as our treatment variable and estimate versions of the following model:

$$WikiOutcome_{cb} = \sum_{a=-50}^{50} \beta_a 1(HS)_{cab} + \delta_c + \eta_b + \gamma_{b,s} + \varepsilon_{cb} \tag{2}$$

where $WikiOutcome_{cb}$ is an outcome from the Wikidata for people born in city c and year b. The indicator variable $1(HS)_{ca}$ tracks the ages surrounding the first high school entry in place c, and δ_c, η_b, and $\gamma_{b,s}$ are place, birth year, and state-birth year fixed effects (respectively).

The coefficients of interest are the vector β_a, which track effects across age cohorts. For example, β_{30} is the effect of a high school opening in your birth town at age 30, while β_{12} is the effect of a high school opening in your birth town at age 12. Observing the path of coefficients β_{18} through β_{50} is analogous to looking at pre-trend coefficients in a standard event study. We would not expect

[20] An alternative approach would be to link the Wikidata eminent people to the historical census, which would allow us to estimate identical models as above.

there to be any differential effects of living in a city that built their first high school *after* one could have received a high school degree. The coefficients β_{15} through β_{17} represent partially treated cohorts. A nearby high school entering during those ages can still offer benefits, though students could not have had a traditional, 4-year high school experience before age 18. Coefficients for ages younger than 14 reflect fully treated cohorts: they had a local high school available to them before traditional high schooling ages. As with the enrollment results, we include a significant "post-period" of children with full access to high schools to account for the possibility of effects that grow over time. For example, β_{-50} indicates children who were born 50 years *after* the first high school existed in their birth city. Throughout, standard errors are clustered by city.

Figure 12 shows the results of estimating Equation 2 on both the intensive and extensive margins of eminent people production. The omitted age is 18. Panel (a) shows the results of a Poisson maximum likelihood equation with the count of the number of eminent people as the outcome variable.[21] Panel (b) shows results from a standard ordinary least squares estimation with a zero-one indicator for whether a given cohort-city produced an eminent person tracked in Wikidata as the outcome variable. In both figures, we see generally flat "pre-trends" in outcomes for people who would have turned 18 or older before high schools opened in their cities. The first dashed line from the right indicates partly treated children who would have been 17 years old or younger when their high school opened, and the second dashed line indicates children who were "fully treated" as their local high school opened before age 13. In both panels, we see that relative to children who were aged 18 at the time that high schools opened, later cohorts were more likely to be recorded in Wikidata.

The implied effect sizes in percentage points are large changes from low pre-construction baseline levels. For example, our Poisson estimates can be interpreted as an approximate 50 percent increase in the production of eminent people for children who were fully treated, relative to a mean

[21] There are a significant number of 0 observations in our data (places with no eminent people for a specific birth cohort), so our preferred count specification is the Poisson.

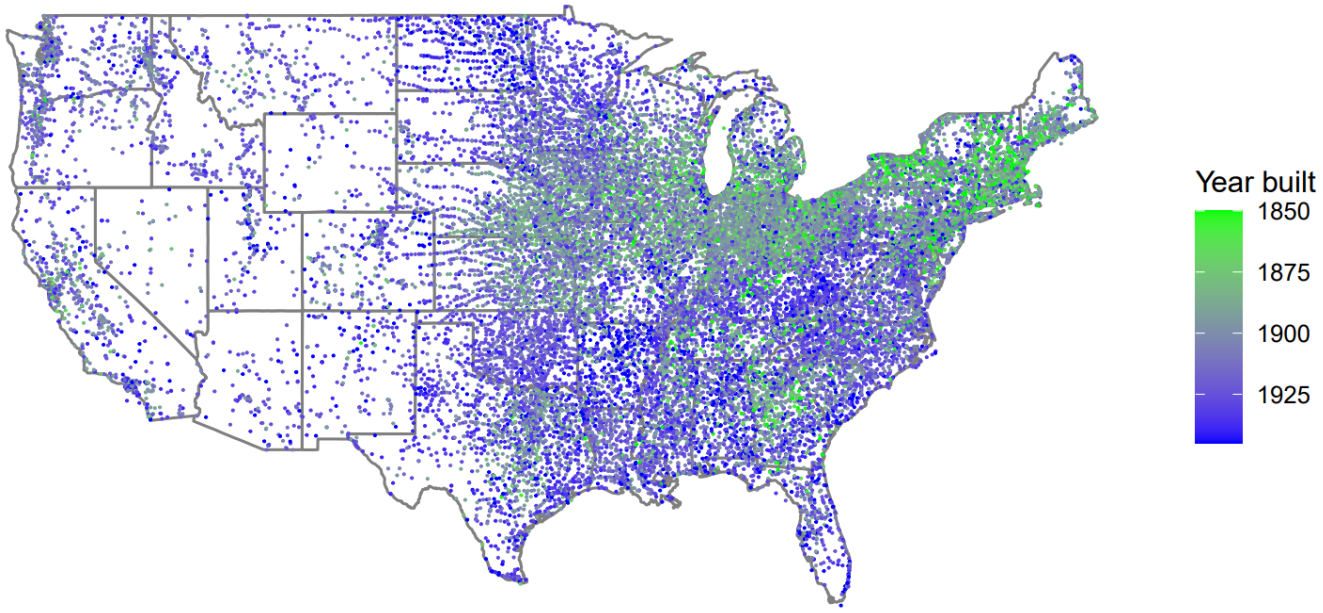
of 0.014 eminent people per birth cohort. The extensive margin effects are less stable over time and grow for later cohorts, reaching a peak of a 2 percentage point increase in the probability of observing an eminent person, relative to a sample mean of 1.4 percentage points.

In Figure 13 we show results separately for men and women. The pattern of results is similar for both men and women, though the percentage effects are larger for women. For example, the extensive margin results for women imply a 250 percent increase in the probability that a town would have an eminent woman born in that town (180% for men), in part owing to the much lower base rate for women. This is also reflected in the Poisson results, though the coefficients for women are less precisely estimated than the effects for men.

5 Conclusion

In this paper, we introduce a new panel covering all high schools established in the United States until 1950 and present evidence on the relationship between the timing of high school entry, student enrollment, and downstream labor market outcomes. Linking children to adult labor market outcomes, we show that high school access boosted labor supply, increased job quality, and reduced the probability of marriage by age 26, with effects concentrated among women. Using structured biographies of eminent adults (e.g., scientists and politicians), we also find that high school access boosted the probability that a child would become eminent, increasing access to the most prestigious occupations in the U.S.

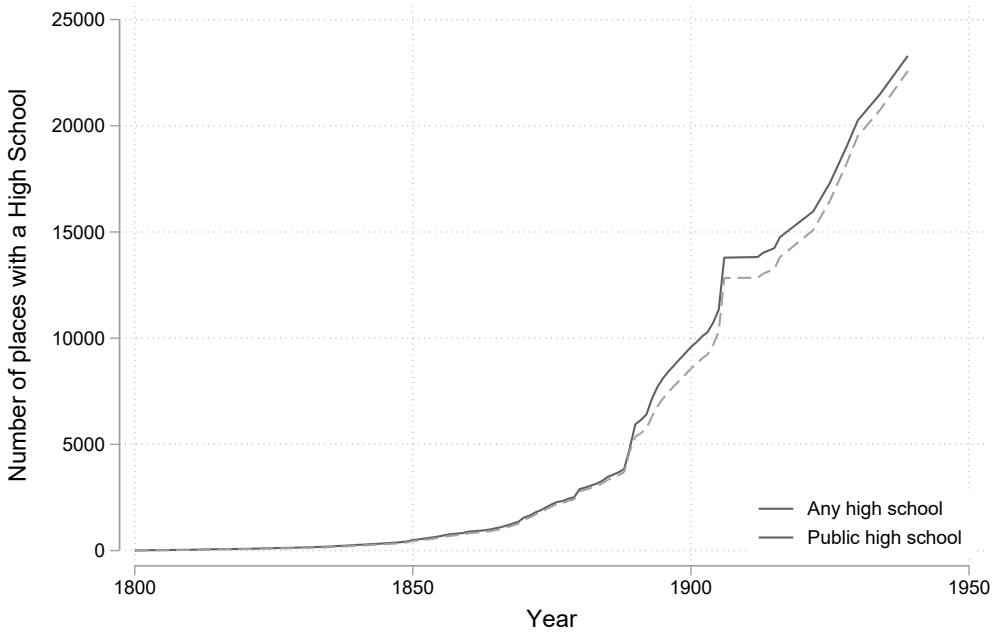
Figure 1: Map of cities, by year of first high school constructed



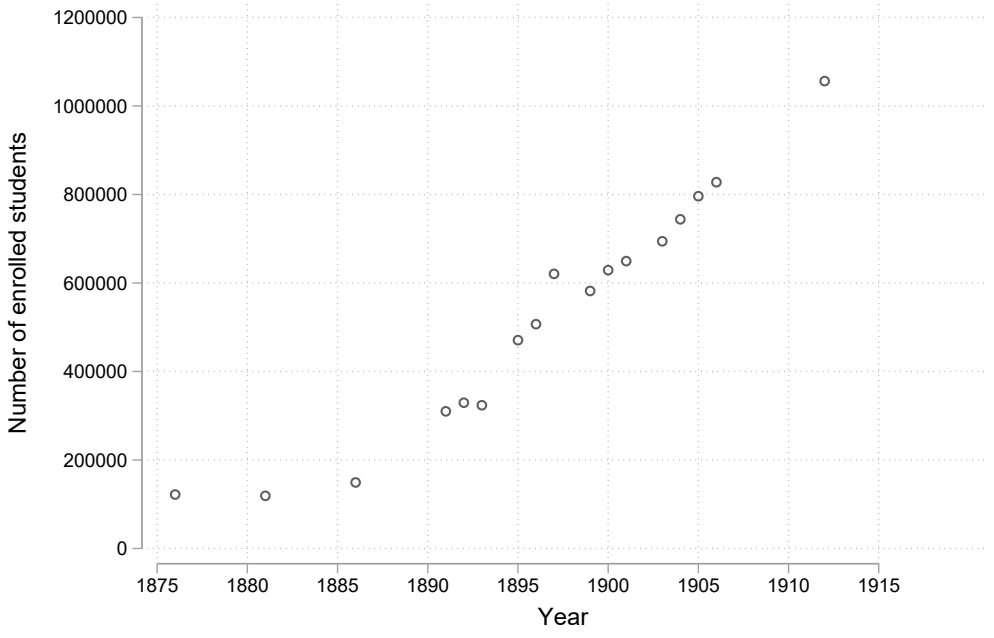
Notes: This map shows the founding dates of the first high school built in each of 26,877 places that we map to a valid longitude-latitude.

Figure 2: Towns with high schools and overall enrollment

(a) Number of towns in the U.S. with high schools, by year



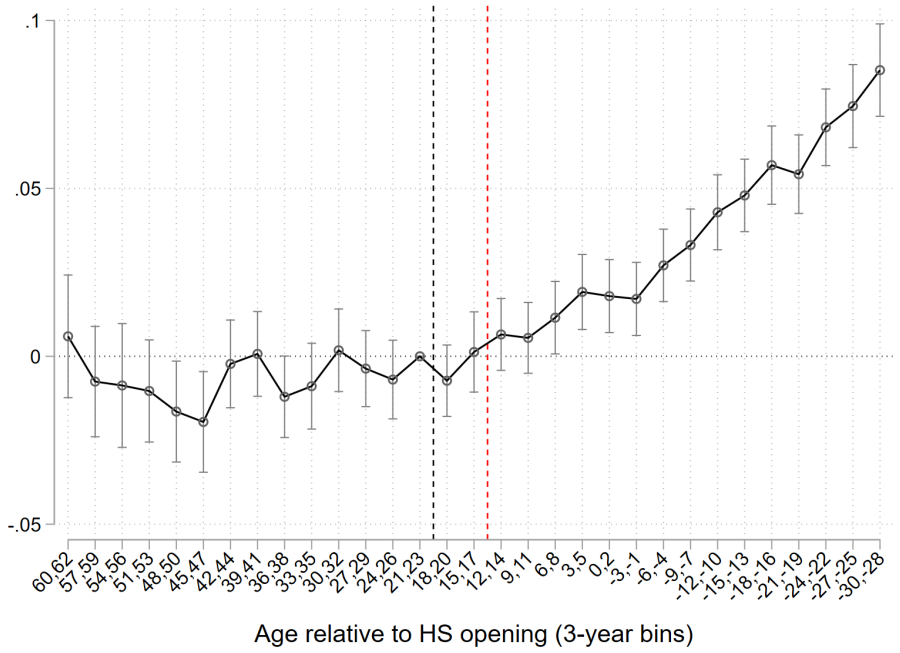
(b) Number of students enrolled in high schools



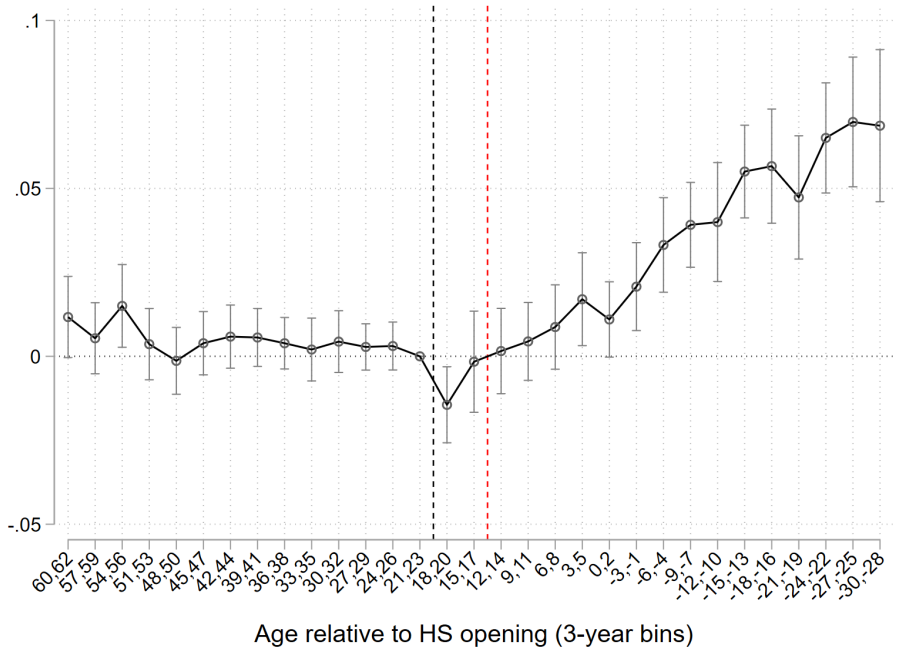
Notes: Panel A shows the number of towns and cities with any high school (the solid line) and with any public high school (the dashed line) from 1800 to 1950. The founding dates are drawn from the panel described in the Data section. Panel B shows the number of students enrolled in high schools in years for which we have valid enrollment data.

Figure 3: Effect of high school entry on 16–18-year-old school enrollment

(a) Two-way fixed effects



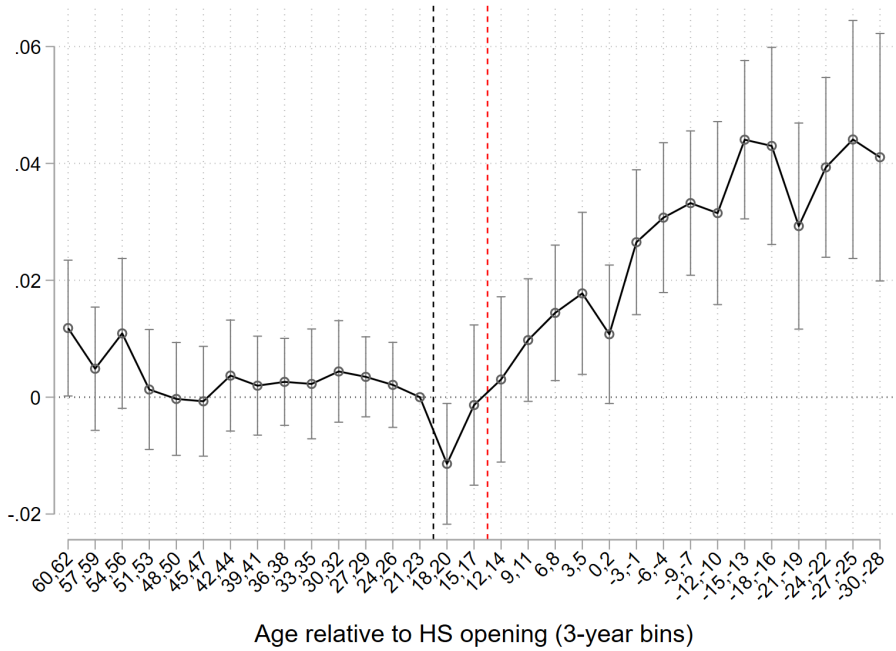
(b) Two-stage Gardner estimation



Notes: This figure shows both our two-way fixed effects and Gardner event study results for school enrollment. Standard errors are clustered by city and each model includes city, birth year, year, and state-year fixed effects.

Figure 4: Effect of high school entry on 16–18-year-old school enrollment, by gender

(a) Women



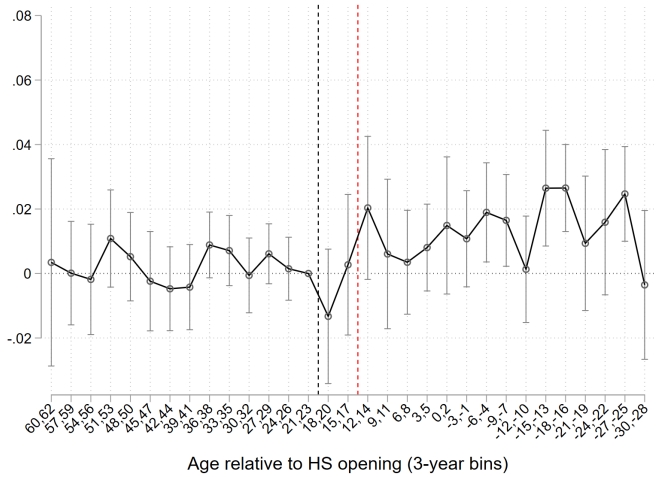
(b) Men



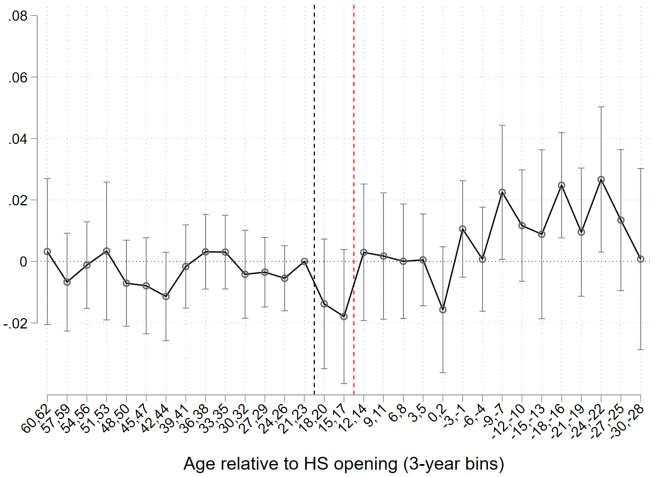
Notes: This figure shows our Gardner event study results for school enrollment by gender. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 5: Effect of high school entry on 13–18-year-old school enrollment, by age

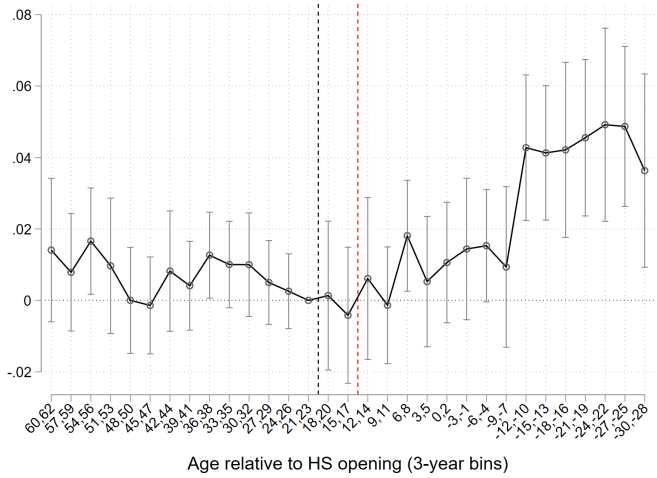
(a) Age 13



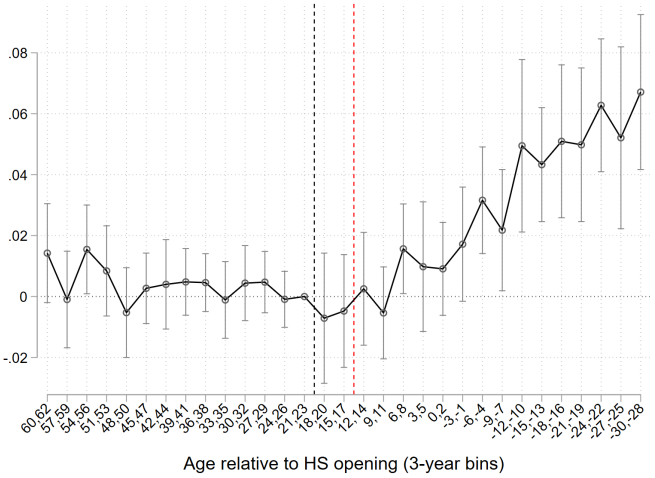
(b) Age 14



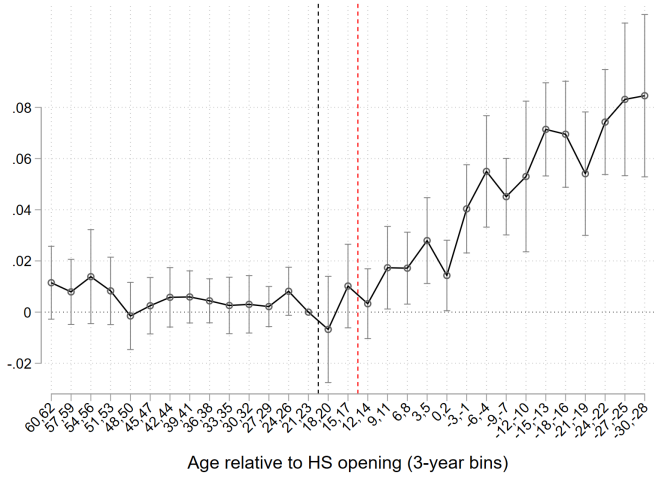
(c) Age 15



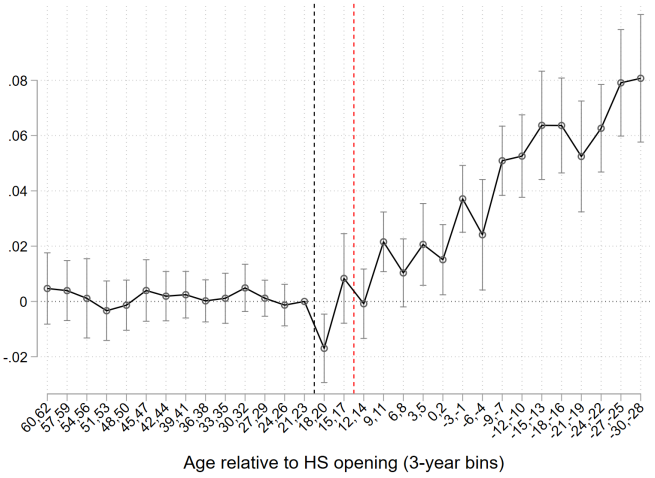
(d) Age 16



(e) Age 17



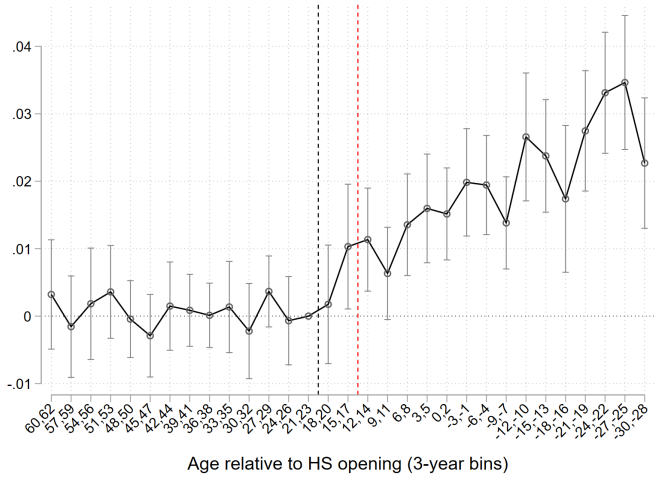
(f) Age 18



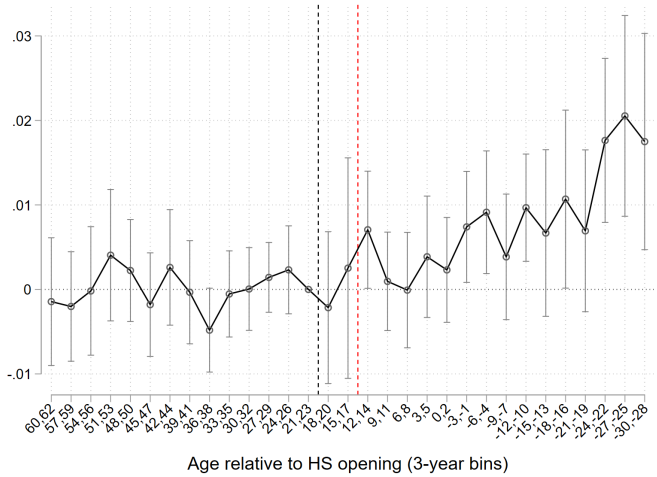
Notes: This figure shows our Gardner event study results for school enrollment by age. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 6: Effect of high school entry on labor force participation

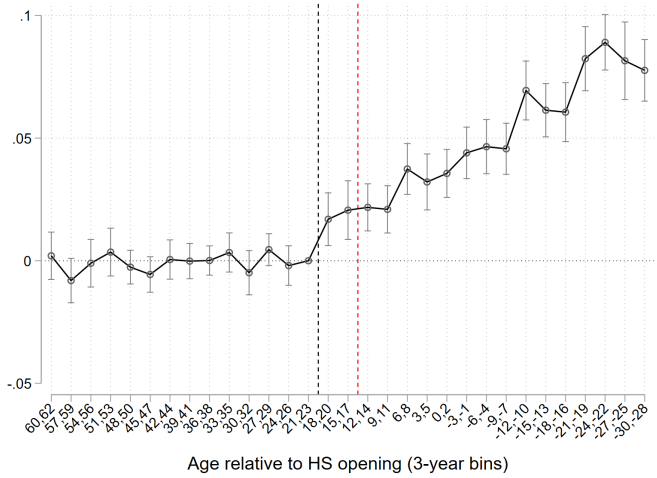
(a) Age 26-28



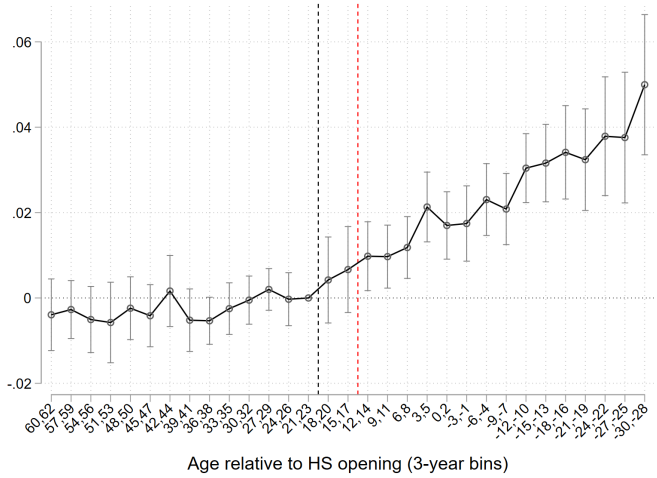
(b) Age 36-38



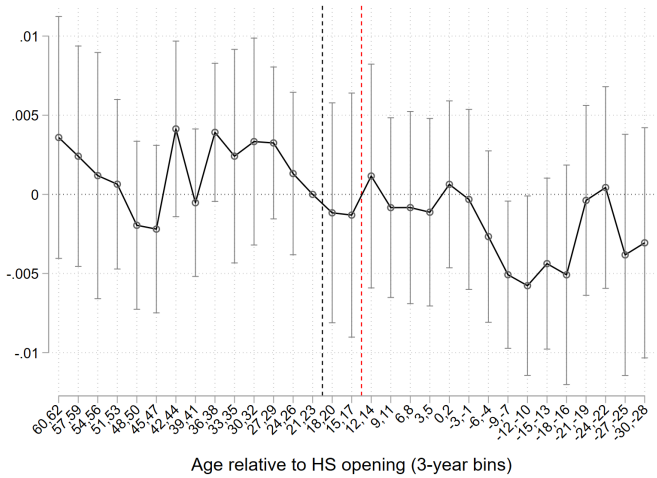
(c) Age 26-28, women



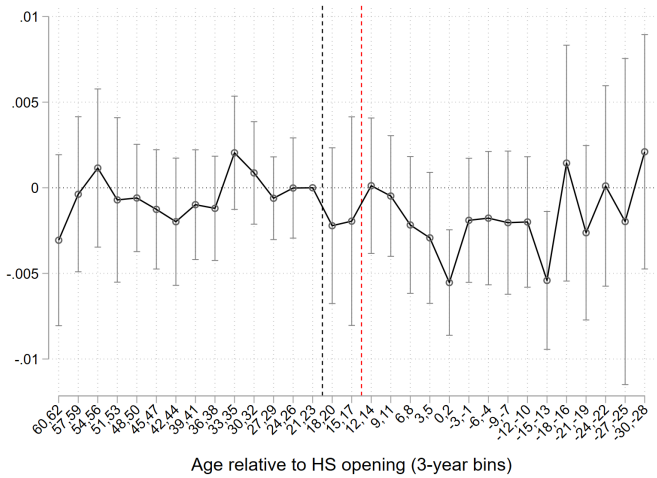
(d) Age 36-38, women



(e) Age 26-28, men



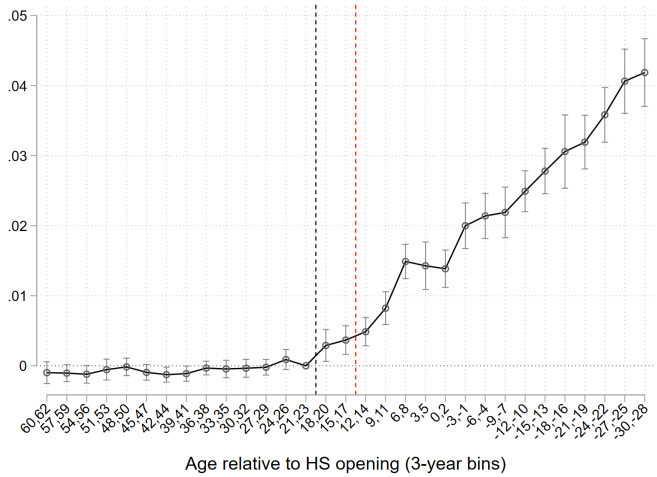
(f) Age 36-38, men



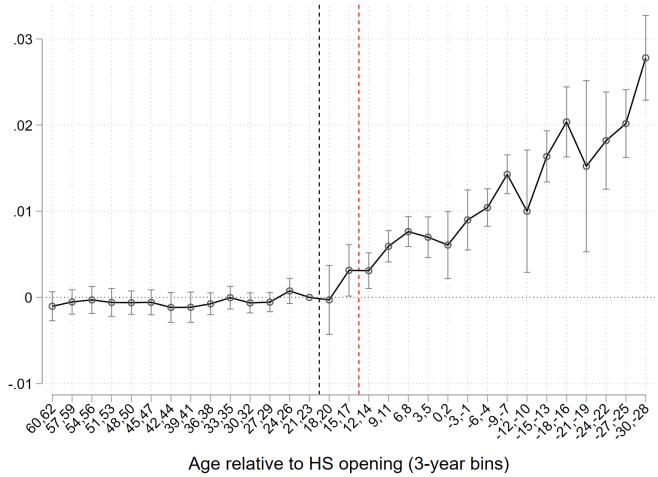
Notes: This figure shows our Gardner event study results for labor force participation. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 7: Effect of high school entry on probability in a clerical occupation

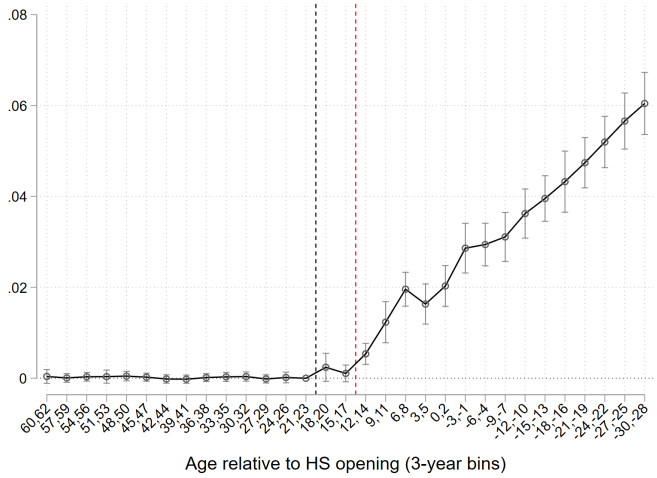
(a) Age 26-28



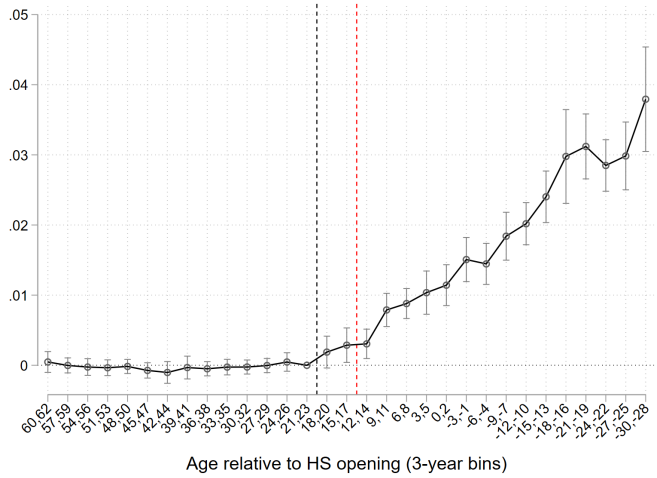
(b) Age 36-38



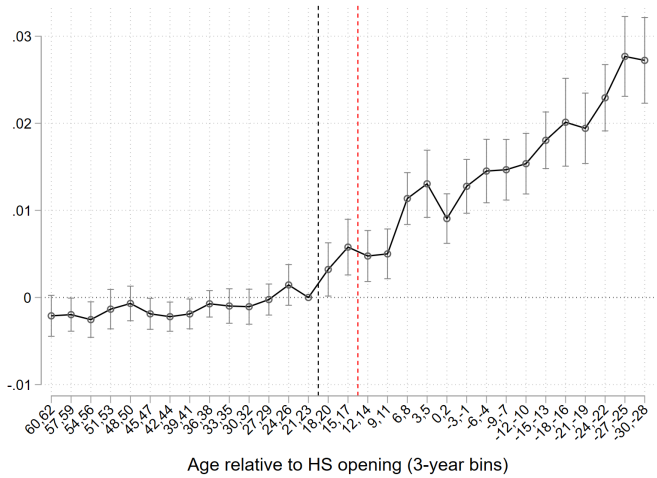
(c) Age 26-28, women



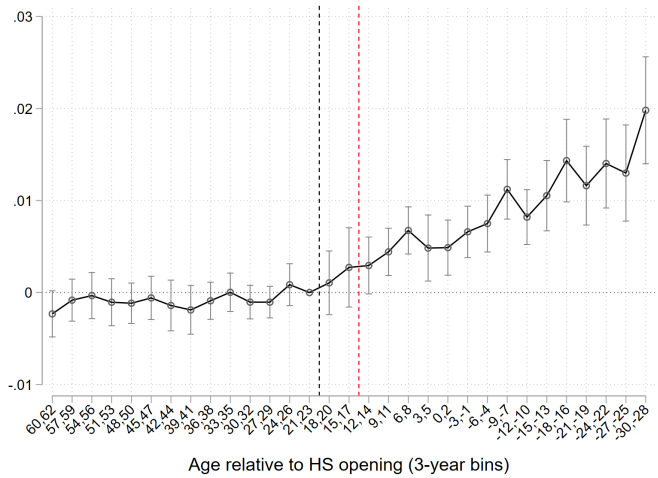
(d) Age 36-38, women



(e) Age 26-28, men



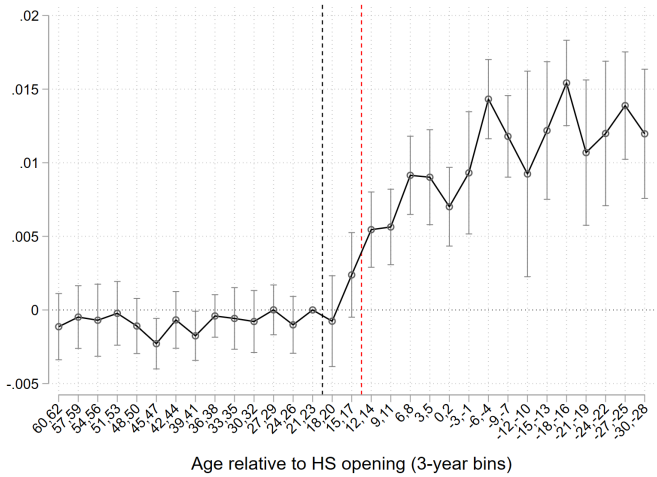
(f) Age 36-38, men



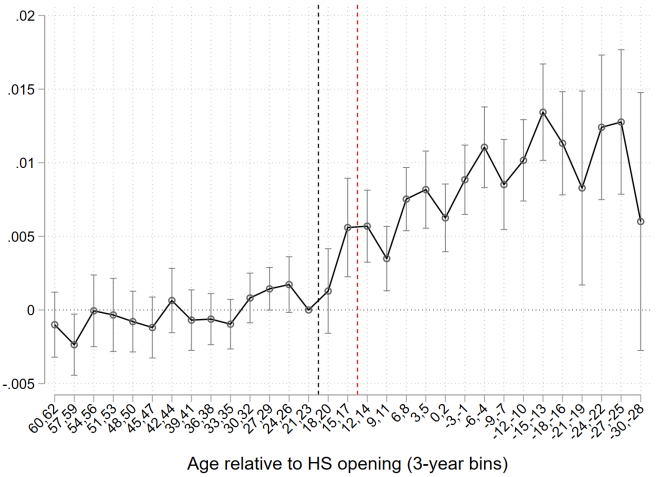
Notes: This figure shows our Gardner event study results for probability employment as clerk. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 8: Effect of high school entry on probability employment in a professional occupation

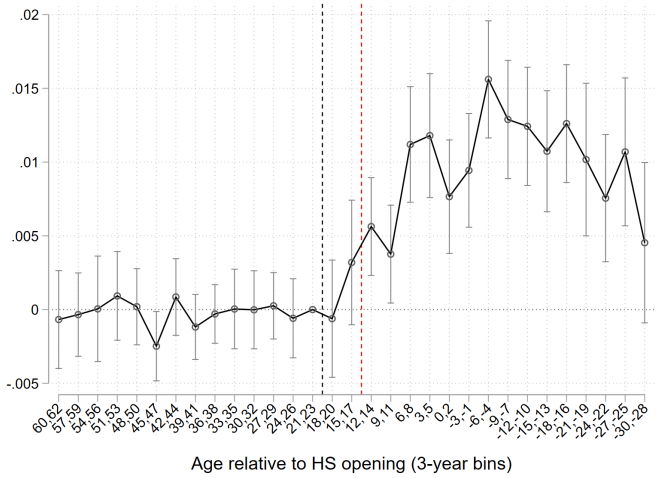
(a) Age 26-28



(b) Age 36-38



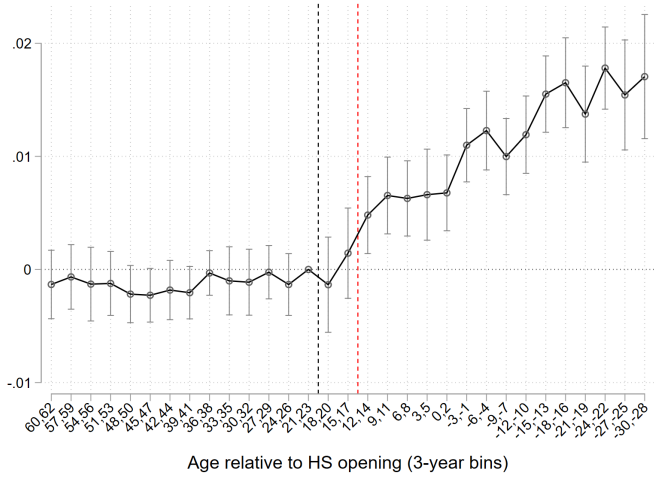
(c) Age 26-28, women



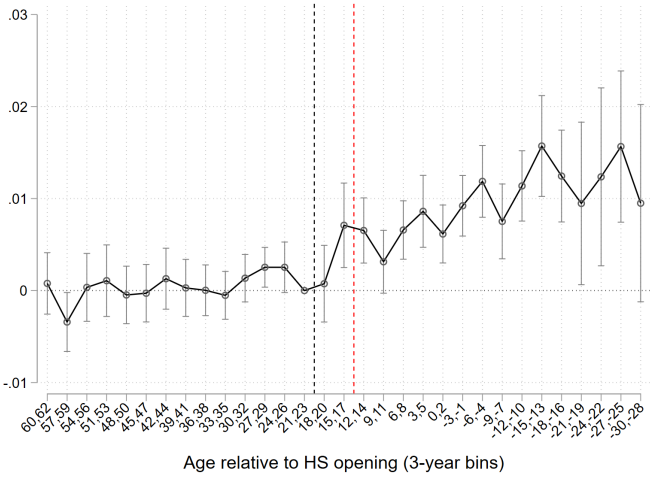
(d) Age 36-38, women



(e) Age 26-28, men



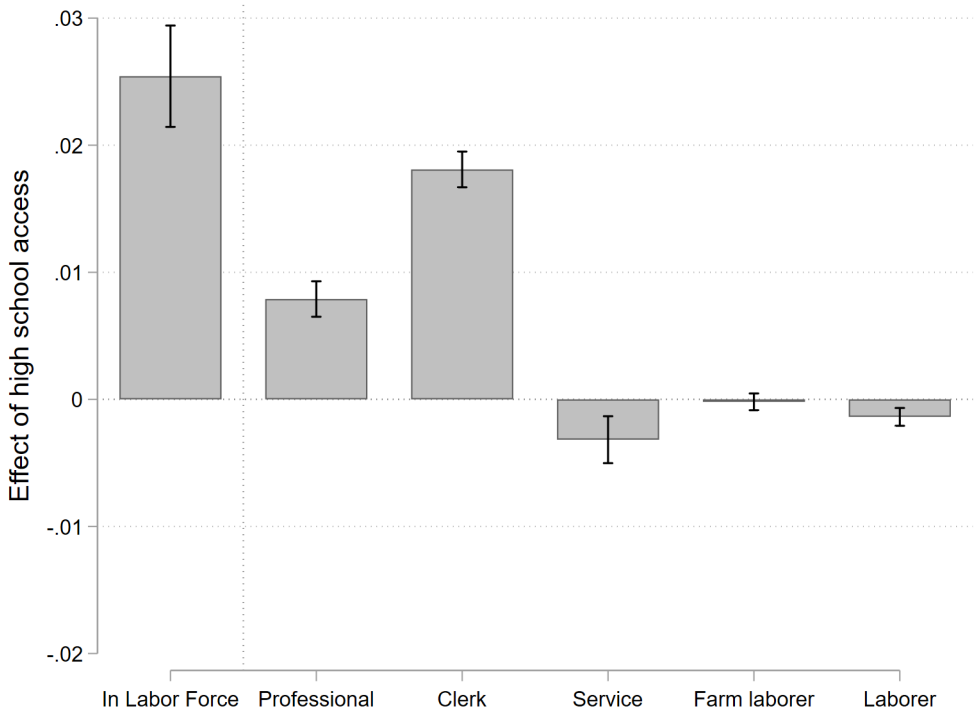
(f) Age 36-38, men



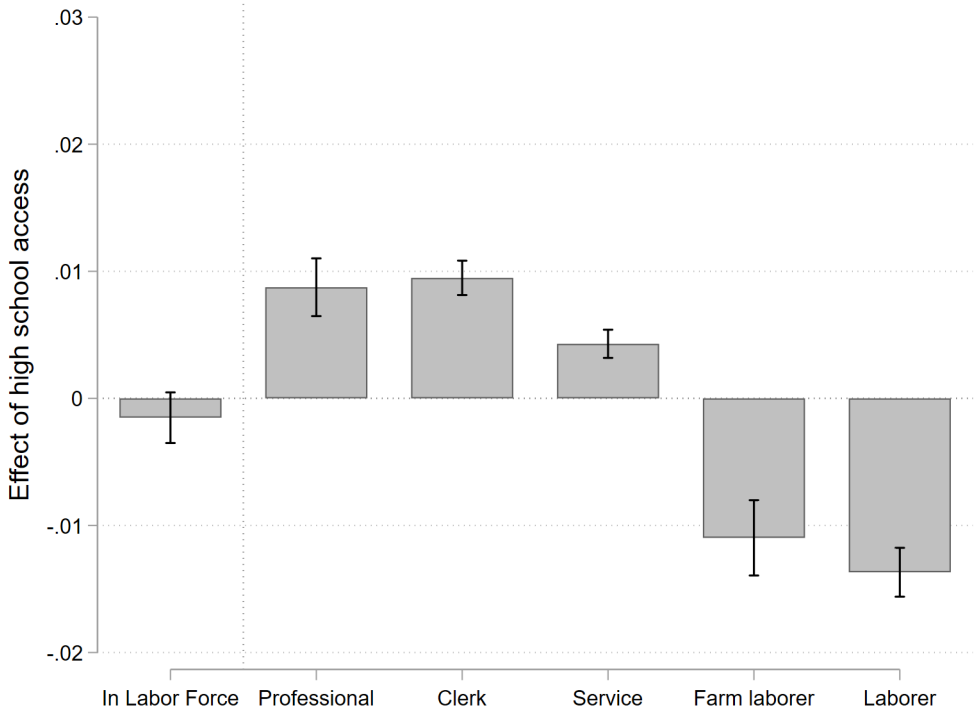
Notes: This figure shows our Gardner event study results for probability employment as a professional. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 9: Age 36-38 occupation and labor force participation results, by gender

(a) Women



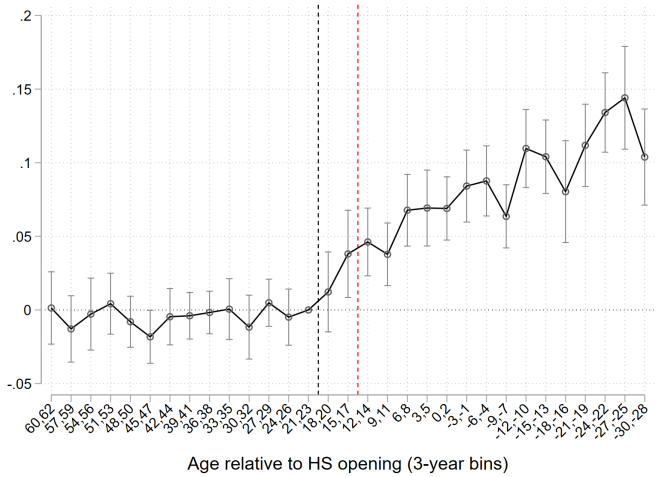
(b) Men



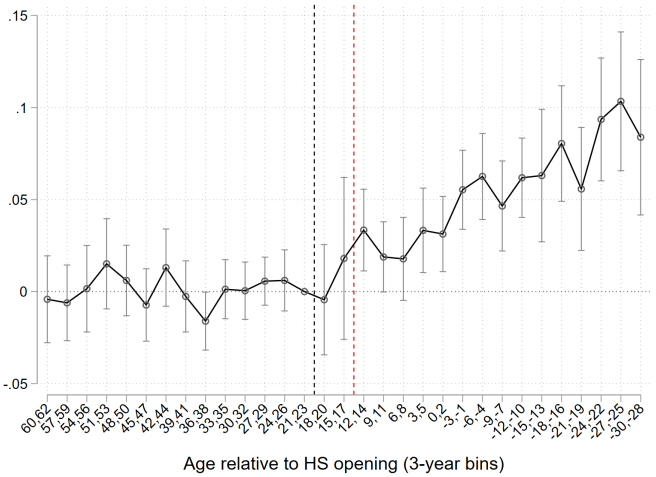
Notes: These figures show the pre-post Gardner difference-in-difference specification results for each indicated outcome and gender. Panel (a) shows results for women. Panel (b) shows results for men.

Figure 10: Effect of high school entry on logged occupation income

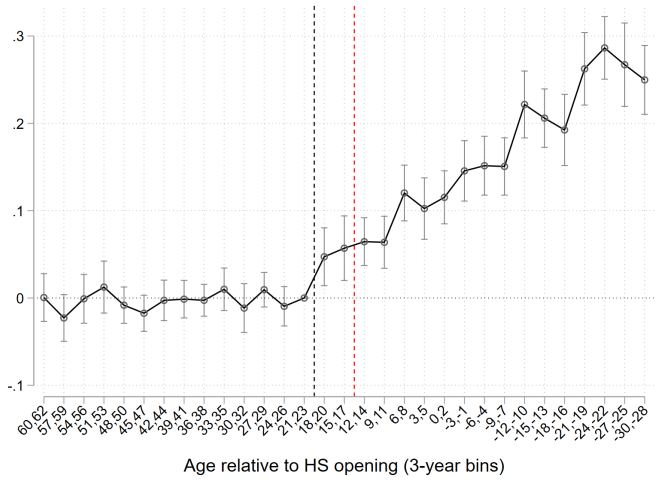
(a) Age 26-28



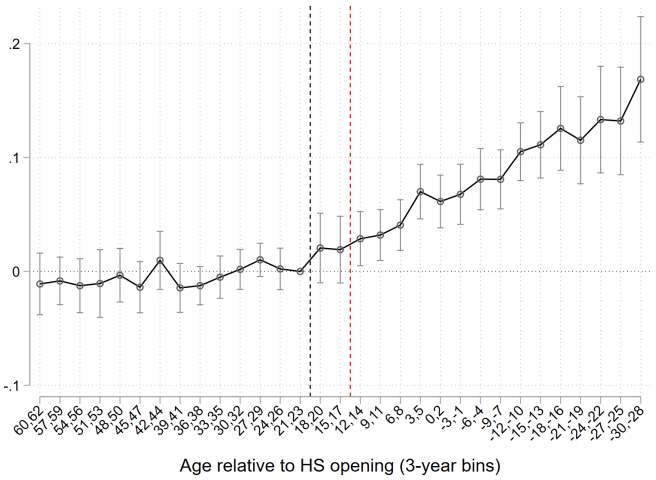
(b) Age 36-38



(c) Age 26-28, women



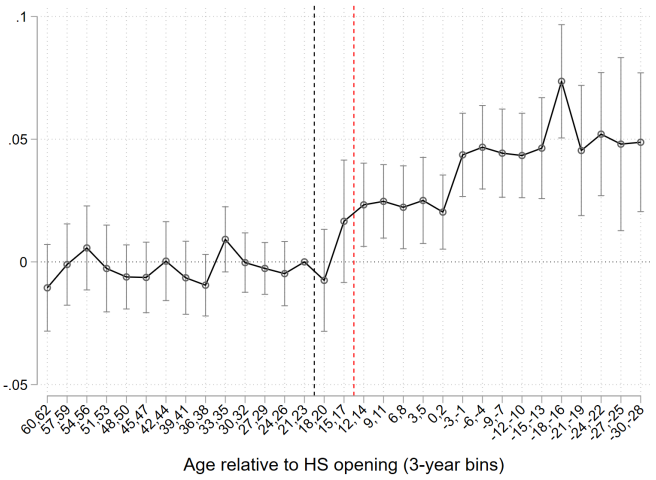
(d) Age 36-38, women



(e) Age 26-28, men



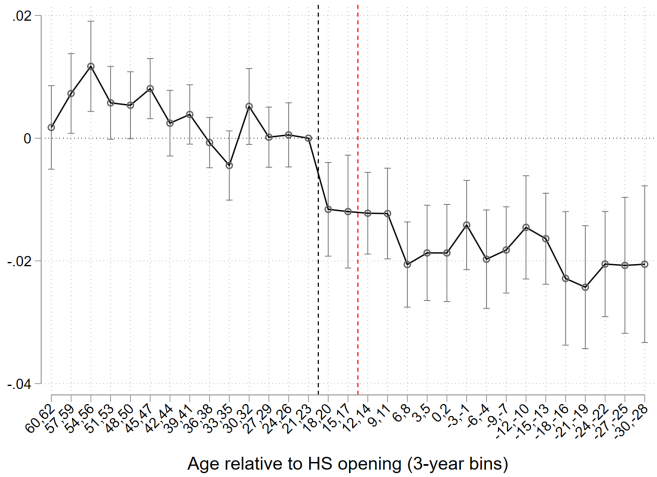
(f) Age 36-38, men



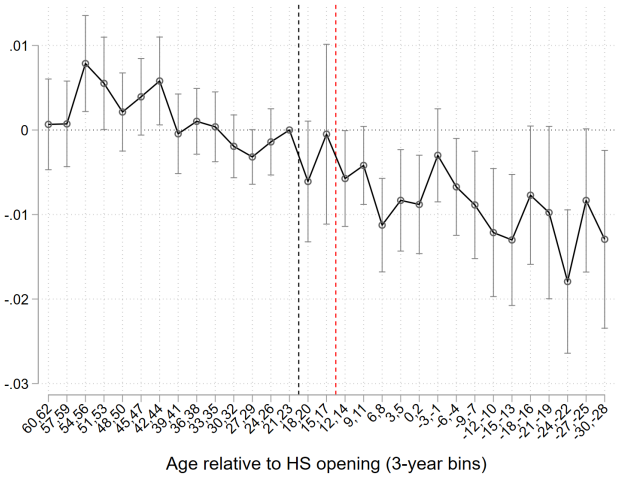
Notes: This figure shows our Gardner event study results for logged imputed occupational income. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 11: Effect of high school entry on marriage probability

(a) Age 26-28



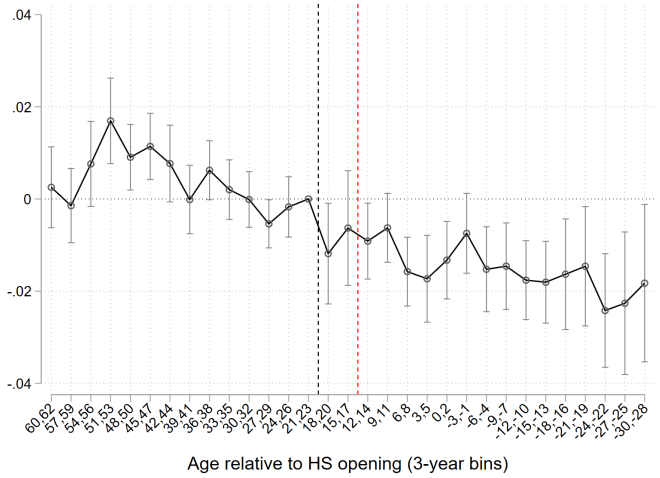
(b) Age 36-38



(c) Age 26-28, women



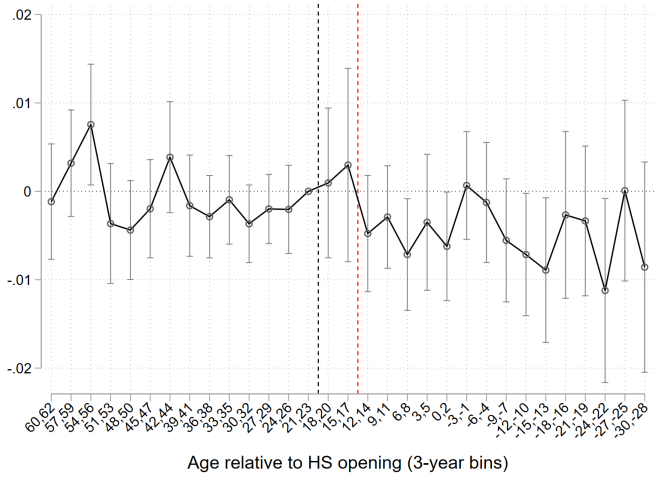
(d) Age 36-38, women



(e) Age 26-28, men



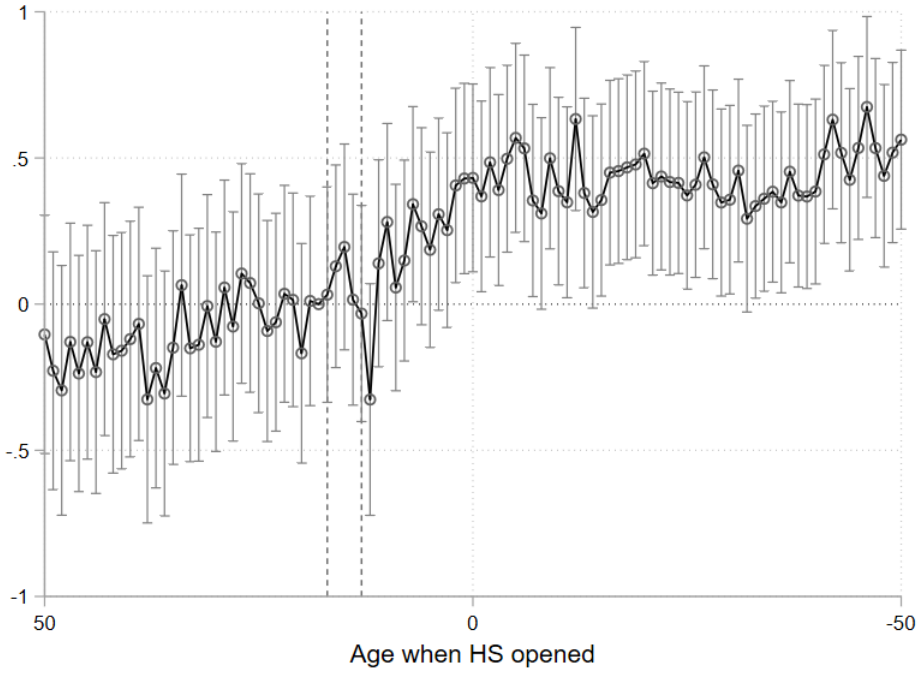
(f) Age 36-38, men



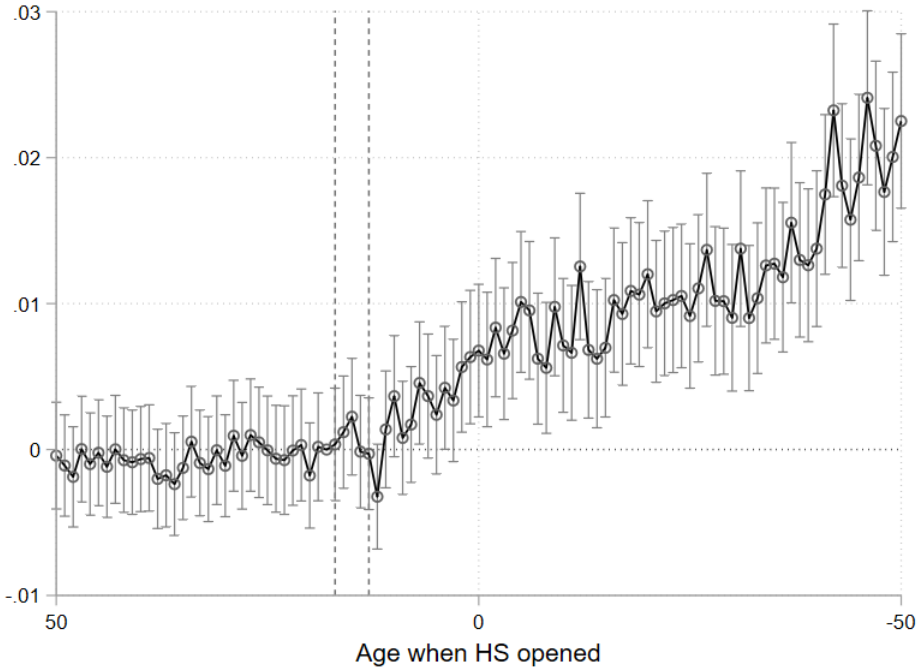
Notes: This figure shows our Gardner event study results for marriage. Standard errors are clustered by city and the model includes city, birth year, year, and state-year fixed effects.

Figure 12: Effect of high school entry on eminent people production

(a) Poisson count (intensive and extensive margin)



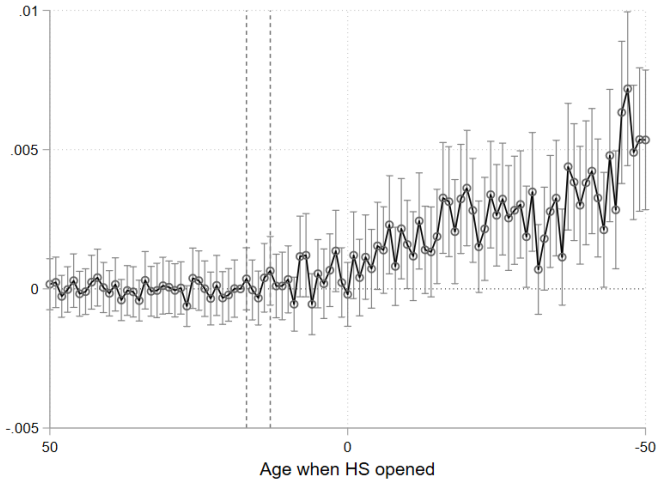
(b) Probability of observing an eminent person (extensive margin)



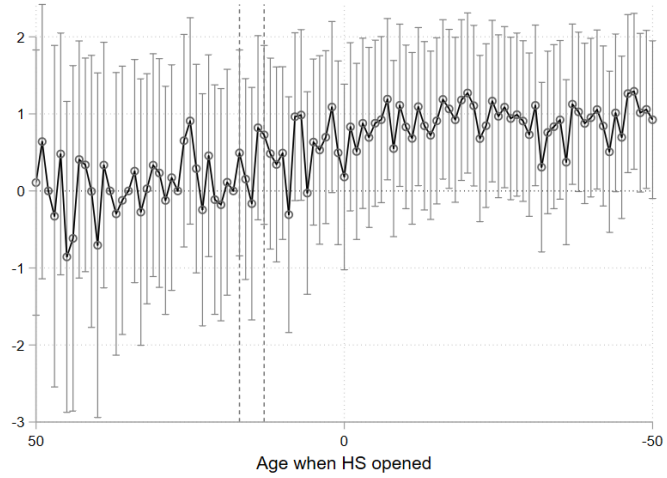
Notes: This shows our baseline event study results for eminent people using Equation 2. Standard errors are clustered by city and each model includes state-year fixed effects.

Figure 13: Effect of high school entry on eminent people production, by gender

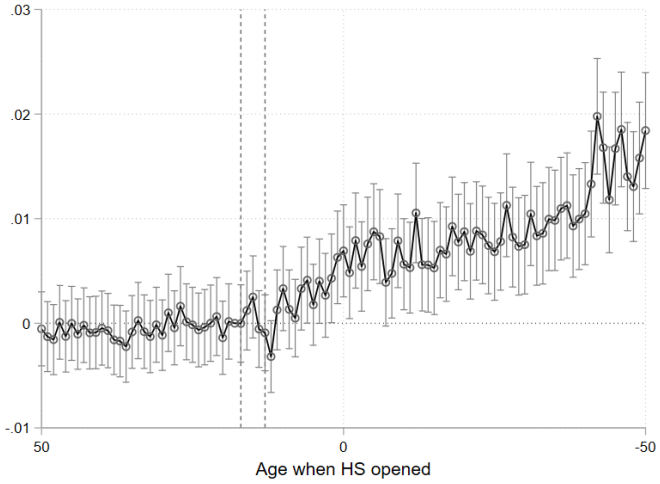
(a) Had female



(b) Count female (Poisson)



(c) Had male



(d) Count male



Notes: This shows our baseline event study results for eminent people using Equation 2 separately by gender. Standard errors are clustered by city and each model includes state-year fixed effects.

References

[1] Aaronson, Daniel and Bhashkar Mazumder. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* Vol. 119, No. 5, pp. 821-888, 2011.

[2] Abramitzky, Ran; Boustan, Leah; Eriksson, Katherine; Rashid, Myera; Pérez, Santiago, 2022, "Census Linking Project: 1850-1860 Crosswalk", https://doi.org/10.7910/DVN/KO5J44, Harvard Dataverse, V2.

[3] Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous. "Innovation and top income inequality." *Econometrica* Vol. 86, No. 1, pp. 1-45, 2019.

[4] Anbinder, Tyler, Dylan Connor, Cormac Ó Gráda, and Simone Wegge. "The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants." *Working paper*, 2021.

[5] Andrews, Michael J. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges." *American Economic Journal: Economic Policy* 15, no. 2 (2023): 1-41.

[6] Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform? Evidence from U.S. Historical Data." *Journal of Economic Literature*, 2020, 58(4): 997-1044.

[7] Baker, Andrew C., David F. Larcker, and Charles CY Wang. "How much should we trust staggered difference-in-differences estimates?." *Journal of Financial Economics* 144, no. 2 (2022): 370-395.

[8] Bell, Alex and Chetty, Raj and Jaravel, Xavier and Petkova, Neviana and Van Reenen, John. "Who becomes an inventor in America? The importance of exposure to innovation." *The Quarterly Journal of Economics* Vol. 134, No. 2, pp. 647-713, 2019.

[9] Berkes, Enrico, Ezra Karger, and Peter Nencka. "The census place project: A method for geolocating unstructured place names." *Explorations in Economic History* 87 (2023): 101477.

[10] Buckles, Kasey, Adrian Haws, Joseph Price, and Haley Wilbert. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." *Working paper*, 2023.

[11] Card, David, Ciprian Domnisoru, and Lowell Taylor. "The intergenerational transmission of human capital: Evidence from the golden age of upward mobility." Journal of Labor Economics 40, no. S1 (2022): S39-S95.

[12] Cascio, Elizabeth U., and Ethan G. Lewis. "Teacher Salaries and Racial Inequality in Educational Attainment in the Midcentury South." *Journal of Labor Economics* 42, no. S1 (2024): S95-S131.

[13] Duflo, Esther. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American Economic Review* Vol. 91, No. 4, pp. 795-813, 2001.

[14] Duflo, Esther. "The medium run effects of educational expansion: evidence from a large construction program in Indonesia." *Journal of Development Economics* Vol. 74, pp. 163-197, 2004.

[15] Gardner, John. "Two-stage difference-in-differences." *Working paper*, 2021.

[16] Goldin, Claudia. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century." *Journal of Economic History* Vol. 58, No. 2, pp. 345-374, 1998.

[17] Goldin, Claudia. "Egalitarianism and the Returns to Education during the Great Transformation of American Education." *Journal of Political Economy* Vol. 107, No. S6, pp. S65-S94, 1999.

[18] Goldin, Claudia and Kenneth Sokoloff. "The Relative Productivity Hypothesis of Industrialization: The American Case, 1820 to 1850." *Quarterly Journal of Economics* Vol. 99, No. 3, pp. 461-487, 1984.

[19] Goldin, Claudia and Lawrence F Katz. *The Race Between Education and Technology*. Harvard University Press, 2008.

[20] Goldin, Claudia. "The Quiet Revolution that Transformed Women's Employment, Education, and Family." *American Economic Review* 96, no. 2 (2006): 1-21.

[21] Goldin, Claudia. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton University Press, 2021.

[22] Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. *IPUMS Multigenerational Longitudinal Panel: Version 1.0* [dataset]. Minneapolis, MN: IPUMS, 2023.

[23] Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. "The allocation of talent and US economic growth." *Econometrica* Vol. 87, No. 5, pp. 1439-1474, 2019.

[24] Jackson, C. Kirabo, and Claire L. Mackevicius. "What impacts can we expect from school spending policy? Evidence from evaluations in the United States." American Economic Journal: Applied Economics 16, no. 1 (2024): 412-446.

[25] Navarro-Sola, Laia. "Secondary Schools with Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria." *Working paper*, 2021.

[26] Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. *IPUMS Ancestry Full Count Data: Version 3.0* [dataset]. Minneapolis, MN: IPUMS, 2021.

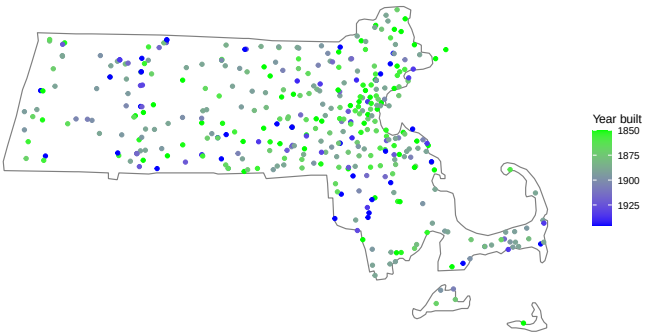
[27] Schmick, Ethan J., and Allison Shertzer. "The Impact of Early Investments in Urban School Systems in the United States" *Working paper*, 2020.

[28] Schmick, Ethan. "The Determinants of Early Investments in Urban School Systems in the United States." *Education Finance and Policy* 19, no. 3 (2024): 409-436.

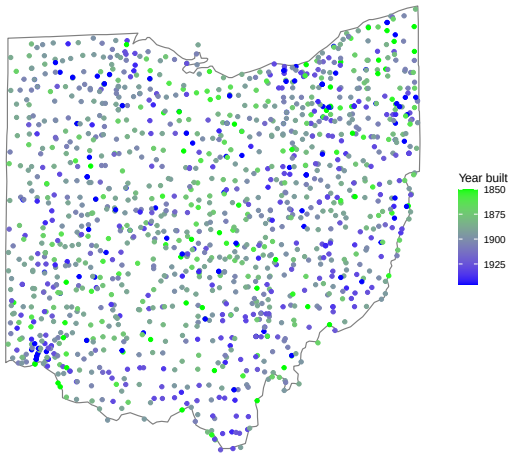
A Figures and tables for online publication

Figure A1: Map of cities, by year of first high school constructed (selected states)

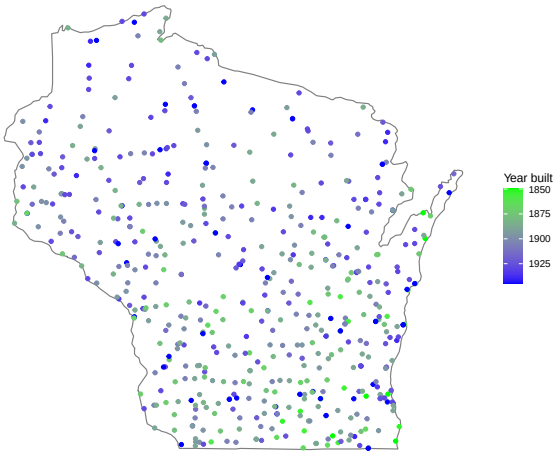
(a) Massachusetts



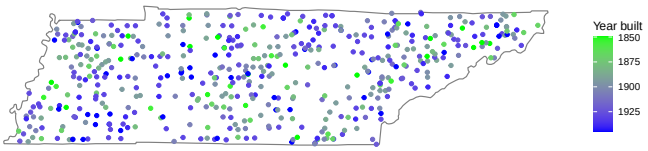
(b) Ohio



(c) Wisconsin

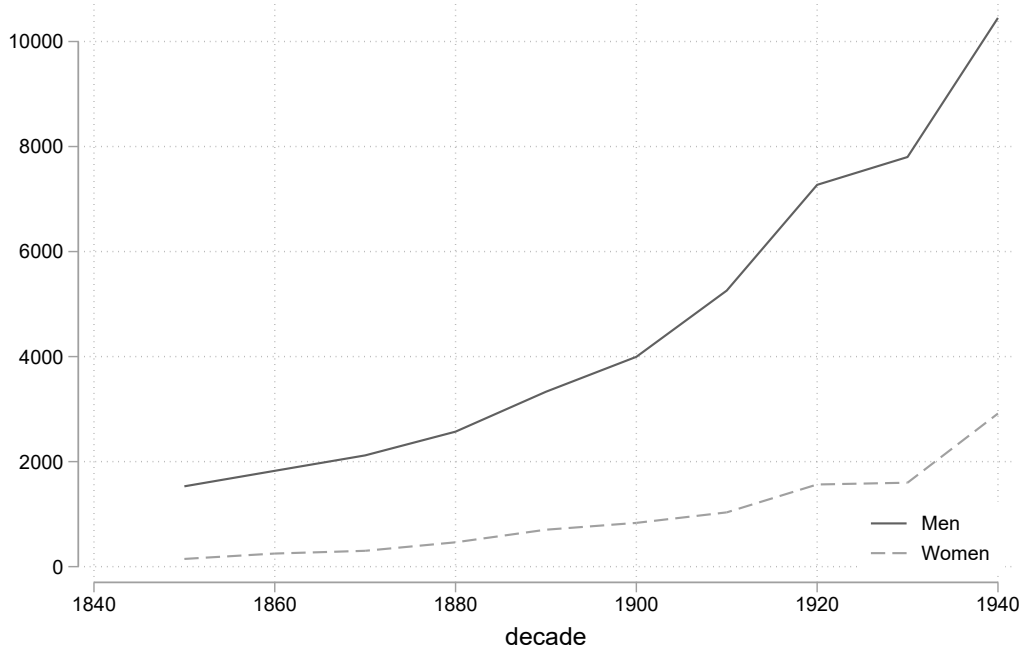


(d) Tennessee



Notes: This map shows the founding dates of the first high school built in each town we can geolocate in Massachusetts (a), Ohio (b), Wisconsin (c), and Tennessee (d).

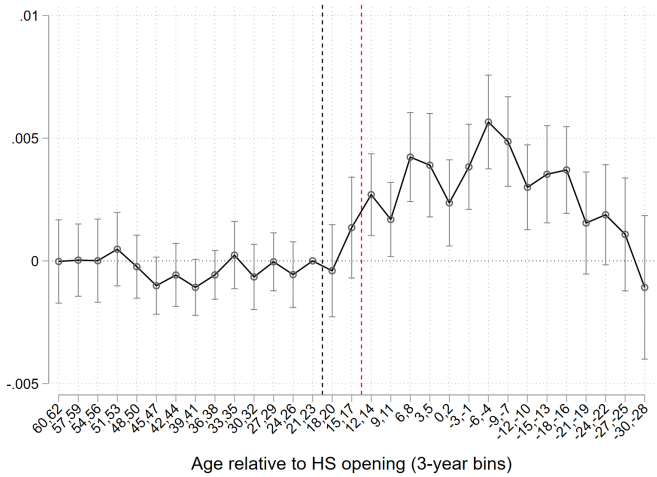
Figure A2: Count of Wikidata eminent people, by birth year and sex



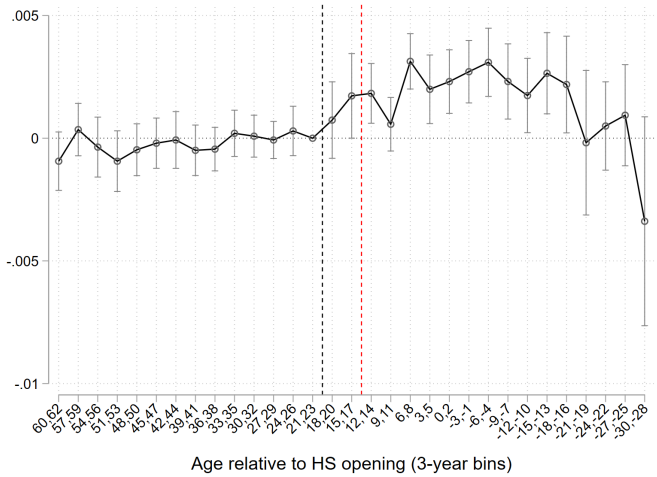
Notes: This plot shows the number of eminent men (solid line) and women (dashed line) in our panel of eminent people from Wikidata. The x-axis represents the birth year of the eminent people and the y-axis is the count of people in our panel.

Figure A3: Effect of high school entry on probability employment as teacher

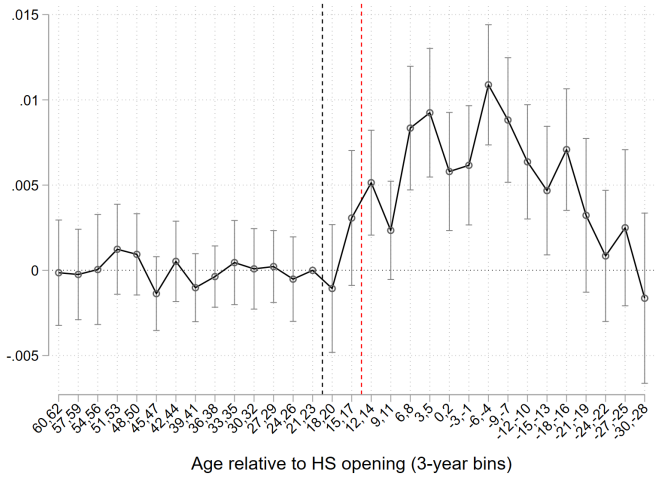
(a) Age 26-28



(b) Age 36-38



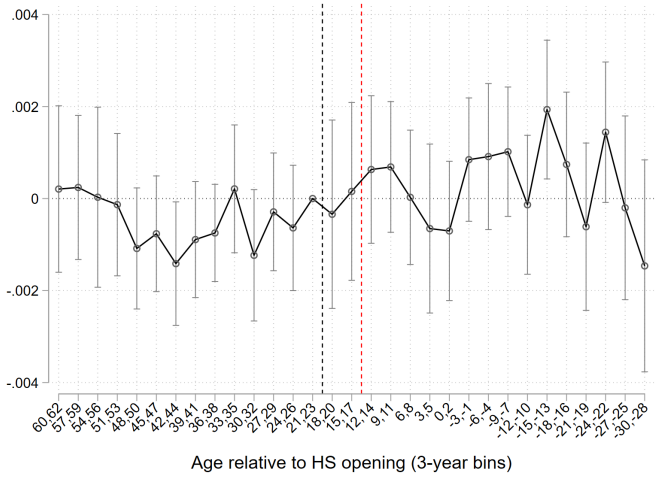
(c) Age 26-28, women



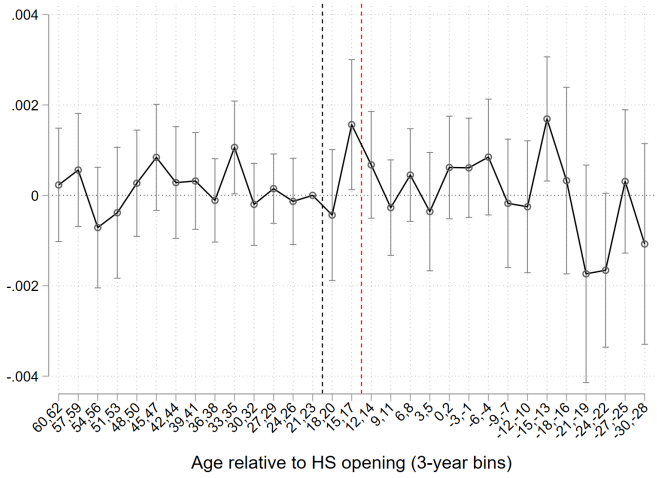
(d) Age 36-38, women



(e) Age 26-28, men



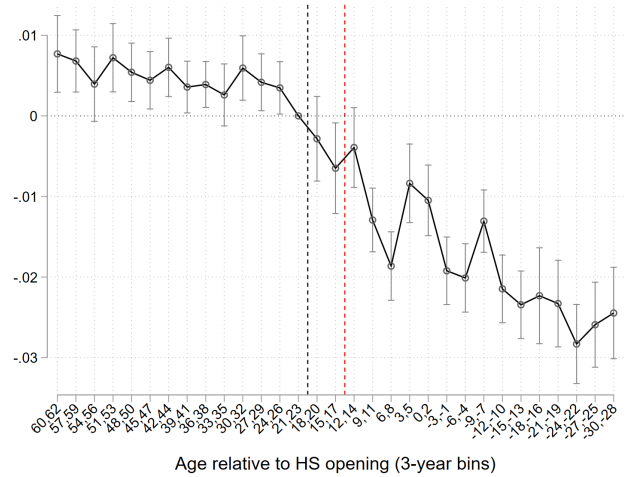
(f) Age 36-38, men



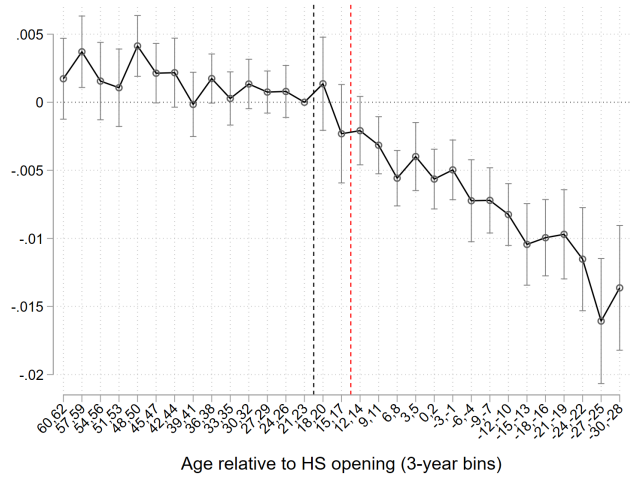
Notes: This figure shows our event study results for each indicated robustness check.

Figure A4: Effect of high school entry on probability employment as farm laborer

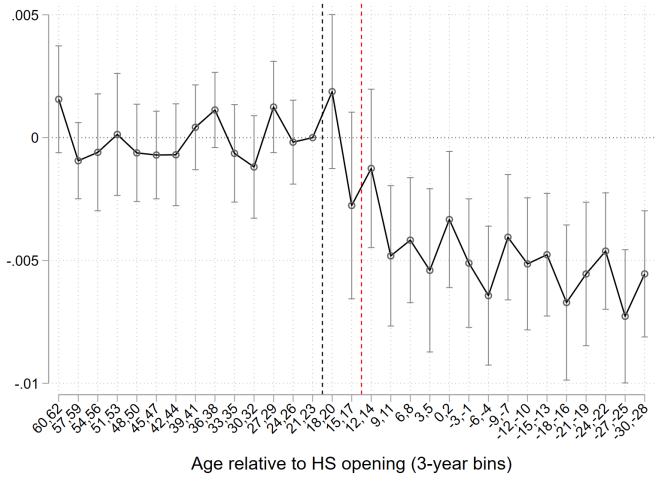
(a) Age 26-28



(b) Age 36-38



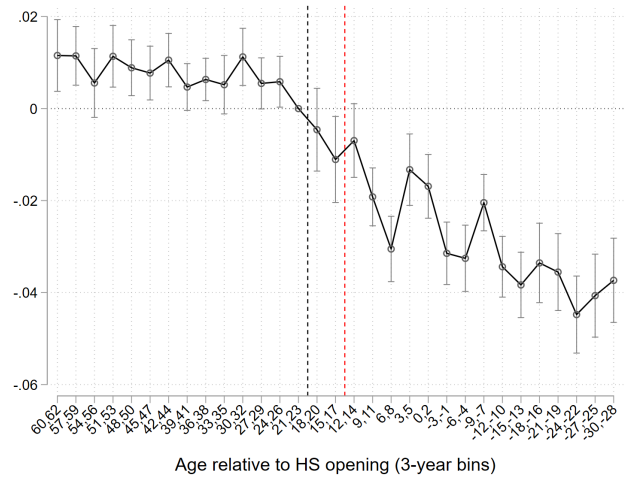
(c) Age 26-28, women



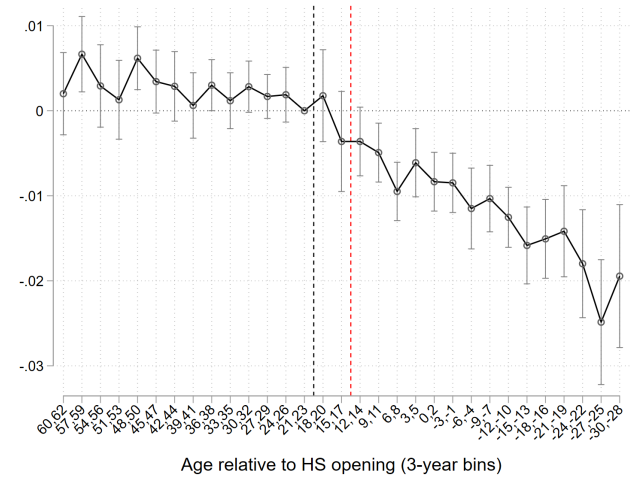
(d) Age 36-38, women



(e) Age 26-28, men



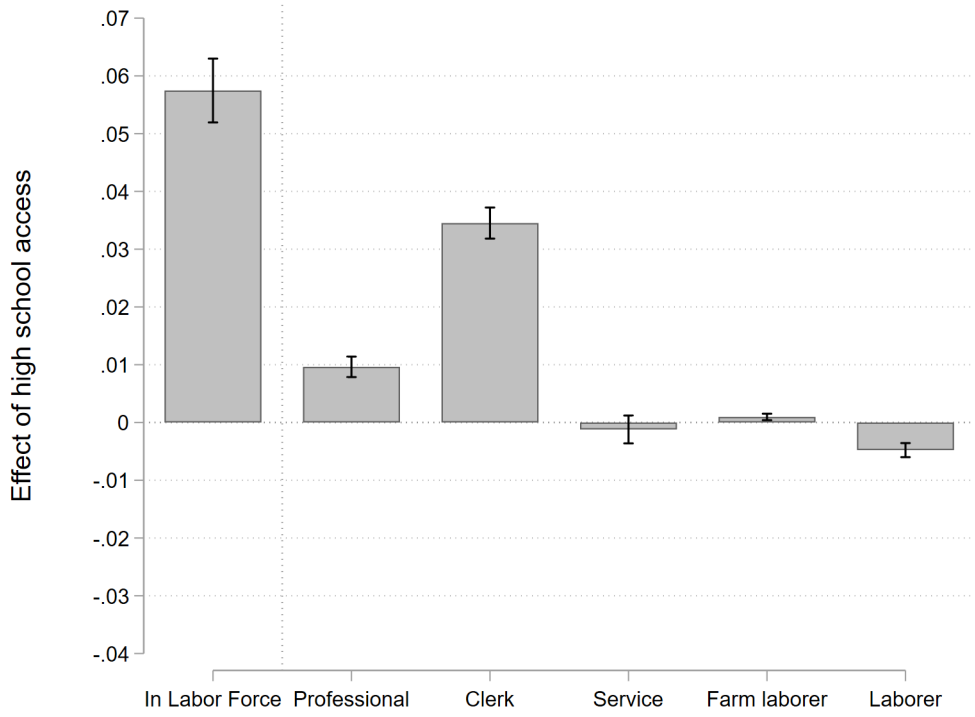
(f) Age 36-38, men



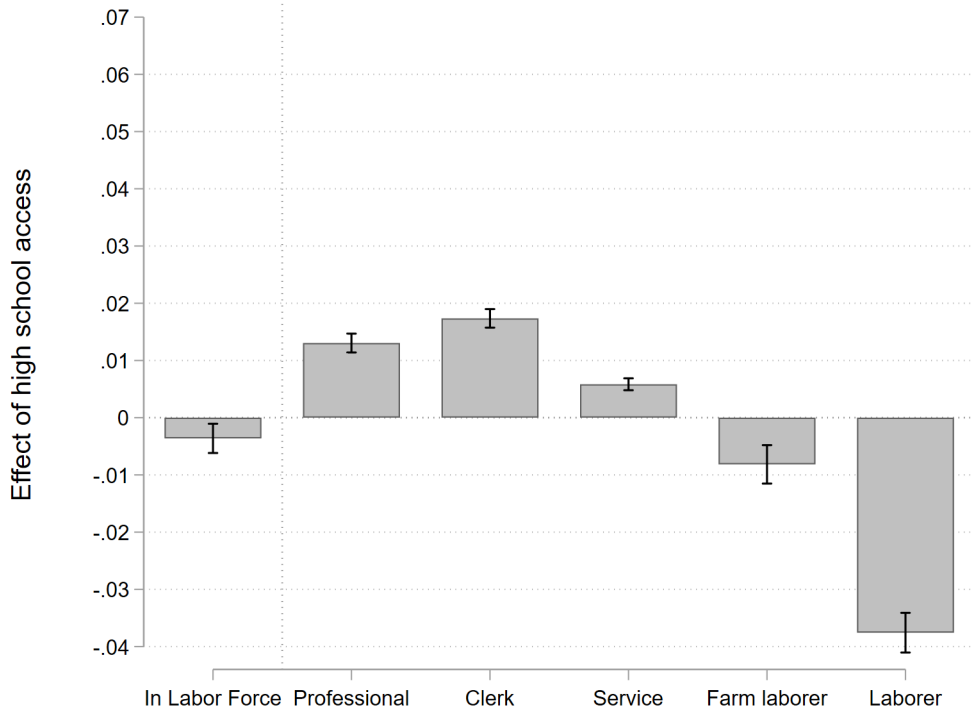
Notes: This figure shows our event study results for each indicated robustness check.

Figure A5: Age 26-28 occupation and labor force participation results, by gender

(a) Women



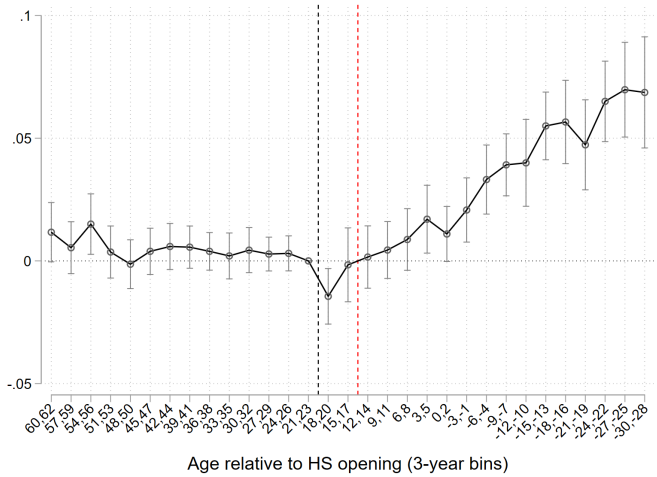
(b) Men



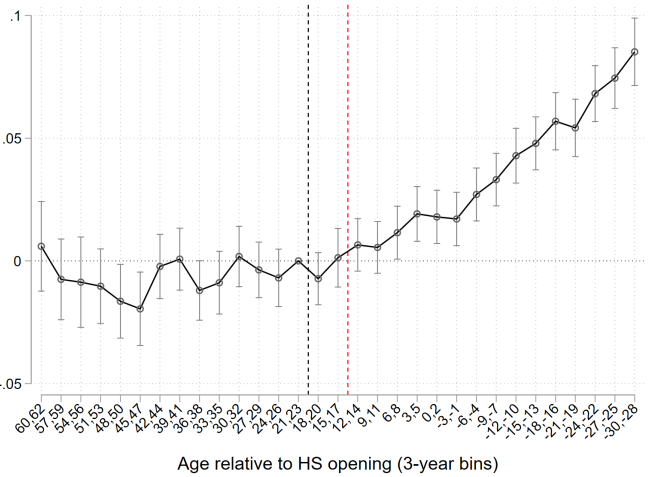
Notes: These figures show the results of a pre-post Gardner difference-in-difference specification for each indicated outcome and gender. Panel (a) shows results for women. Panel (b) shows results for men.

Figure A6: Enrollment results robustness

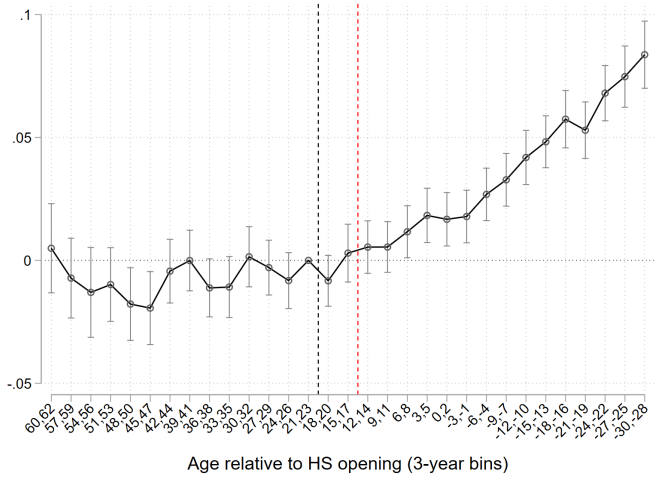
(a) Baseline (Gardner) results



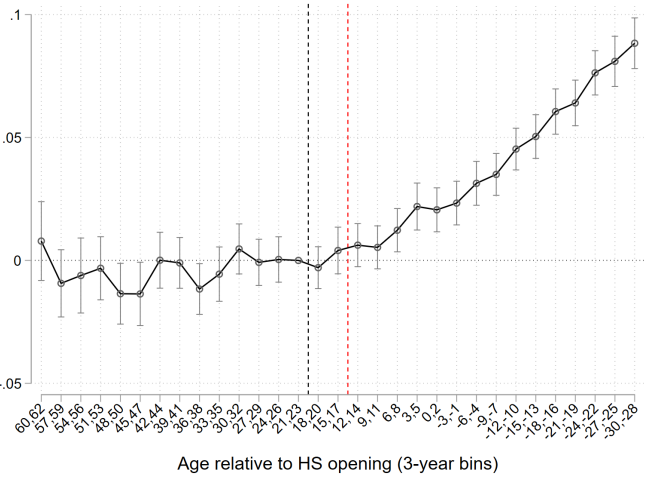
(b) TWFE estimation



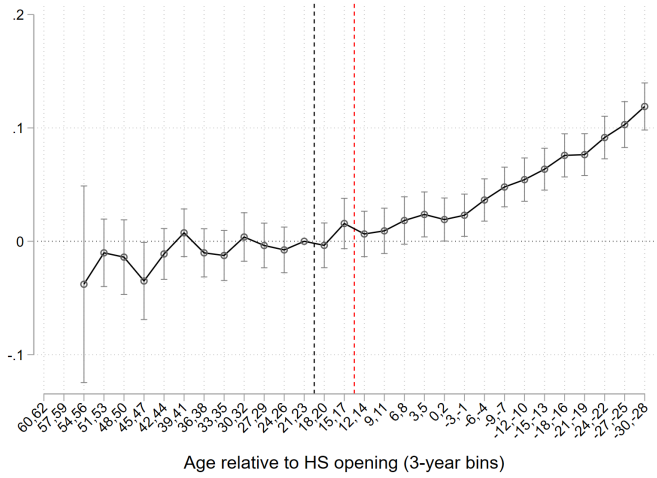
(c) Birth-year-by-state FEs



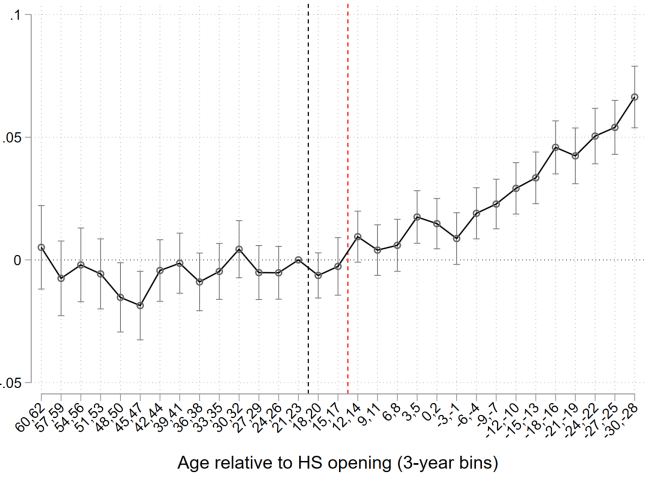
(d) Year-by-county FEs



(e) Non-movers



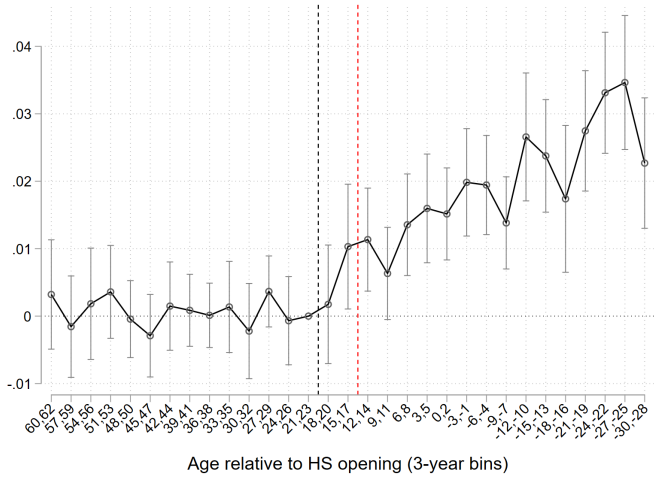
(f) 14-18 year old sample



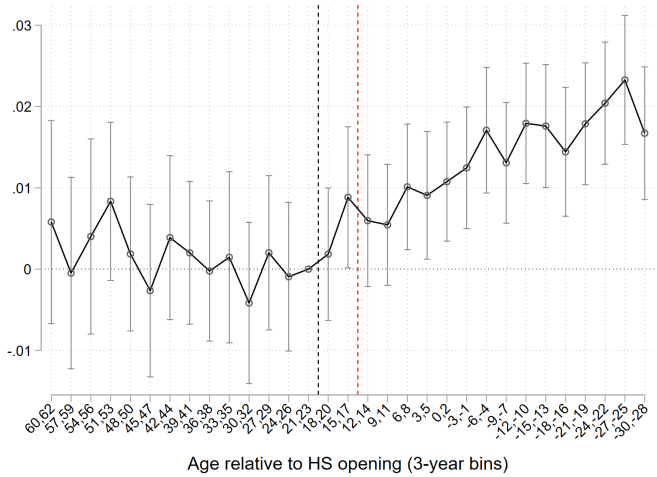
Notes: This figure shows our event study results for each indicated robustness check.

Figure A7: Labor force participation results robustness

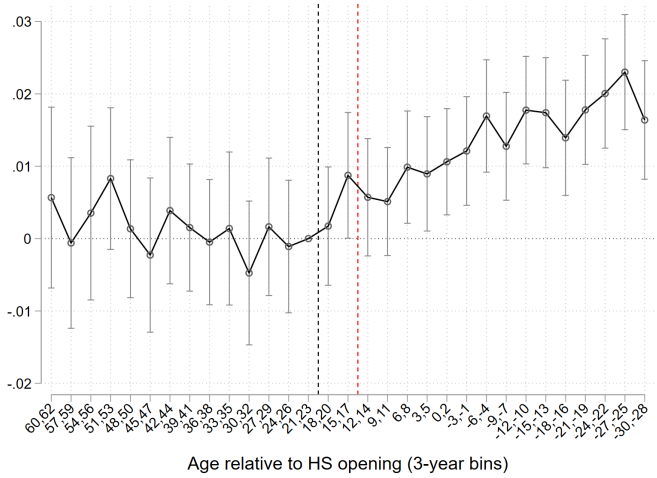
(a) Baseline (Gardner) results



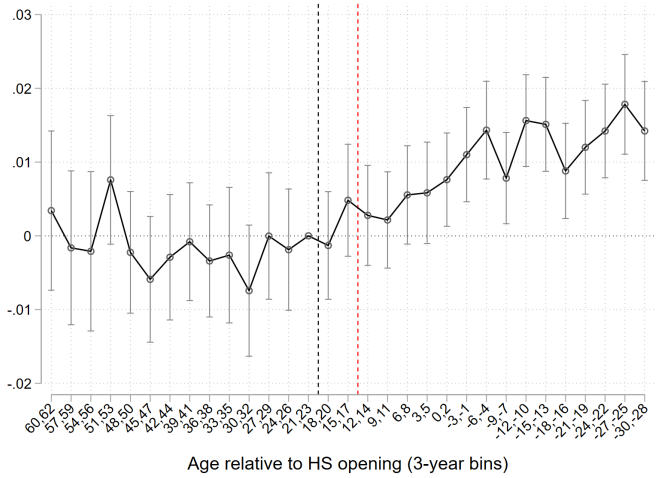
(b) TWFE estimation



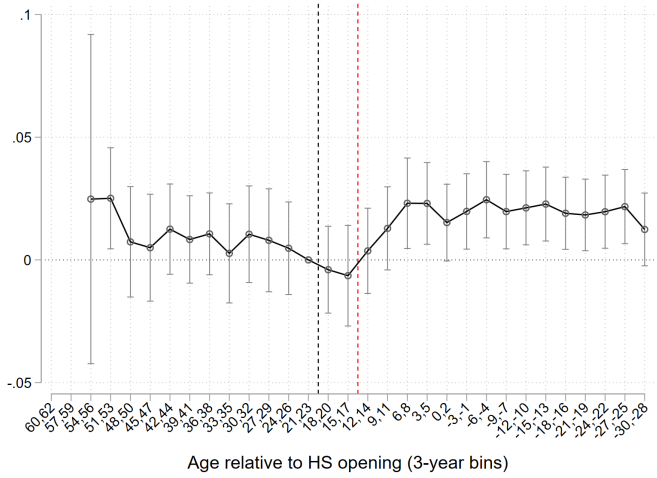
(c) Birth-year-by-state FEs



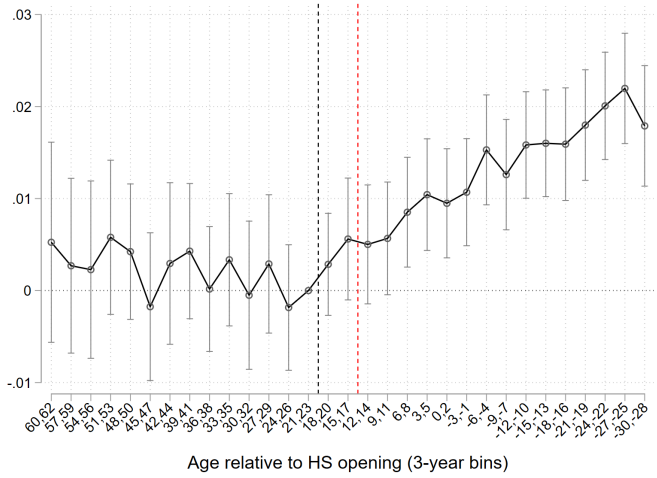
(d) Year-by-county FEs



(e) Non-movers



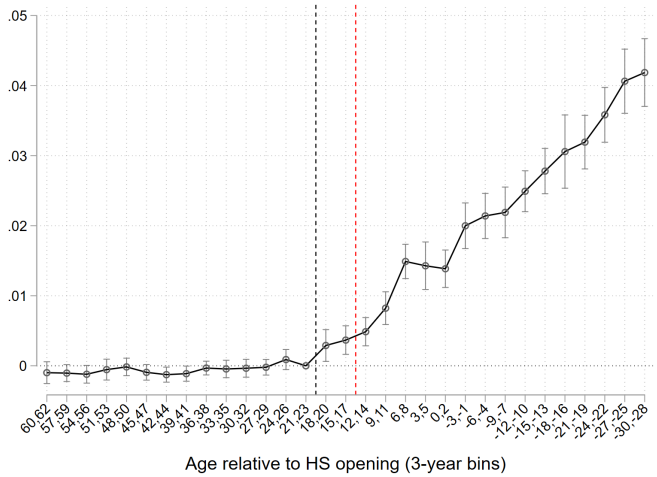
(f) 14-18 year old sample



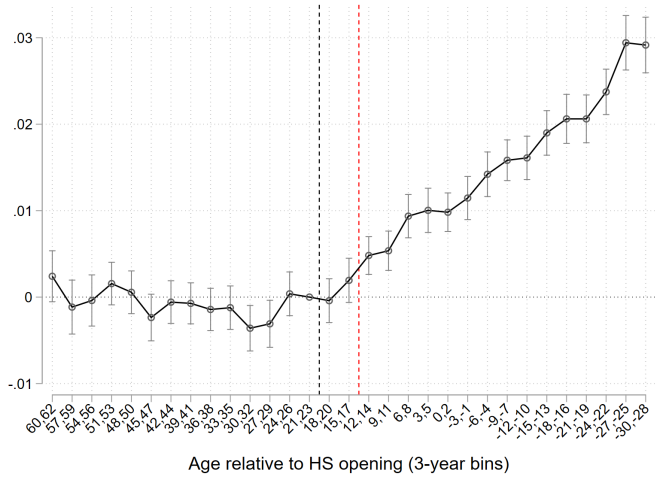
Notes: This figure shows our event study results for each indicated robustness check.

Figure A8: Clerk results robustness

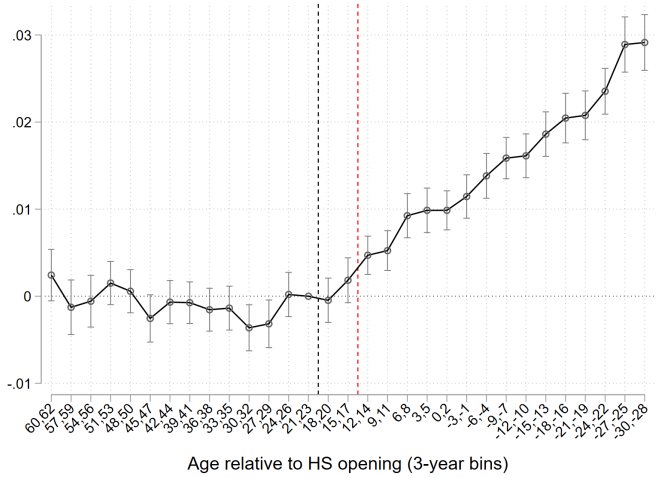
(a) Baseline (Gardner) results



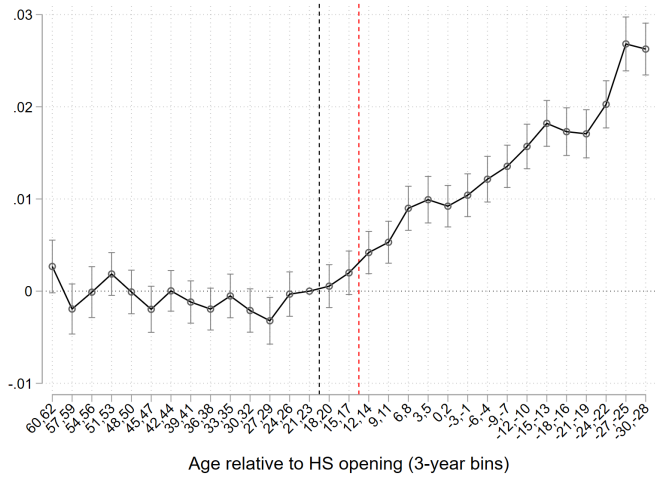
(b) TWFE estimation



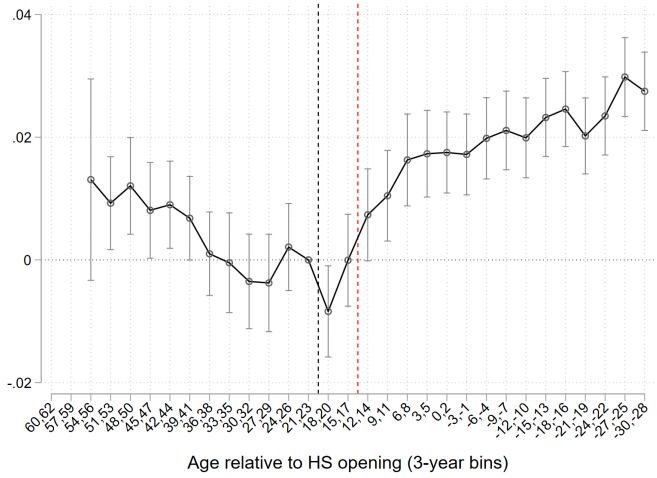
(c) Birth-year-by-state FEs



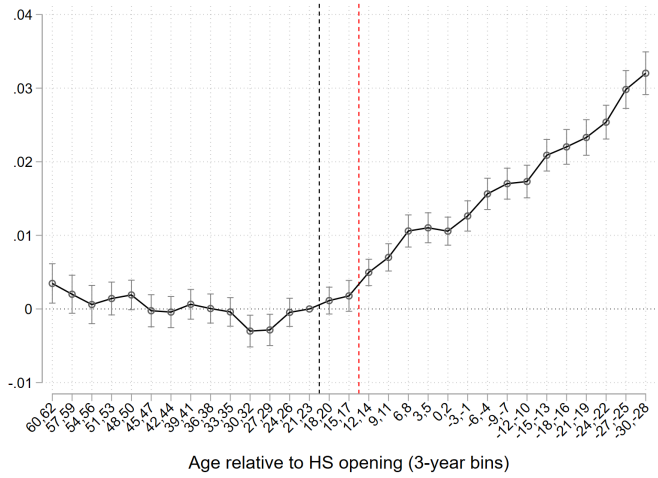
(d) Year-by-county FEs



(e) Non-movers



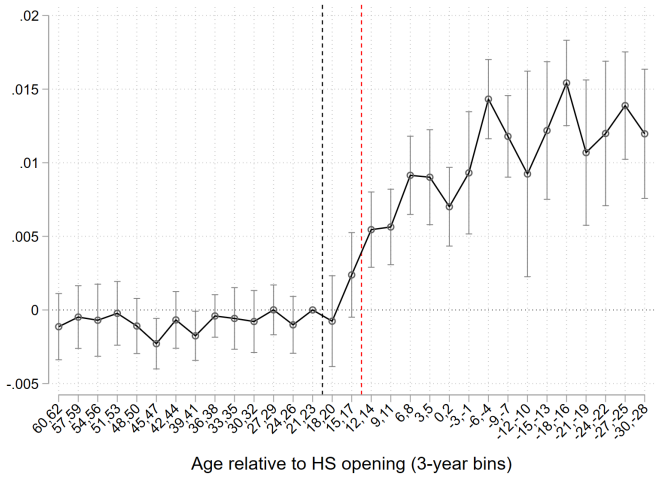
(f) 14-18 year old sample



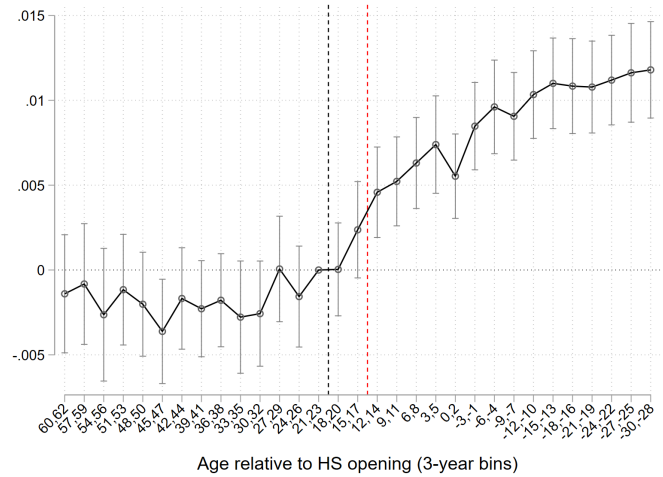
Notes: This figure shows our event study results for each indicated robustness check.

Figure A9: Professional results robustness

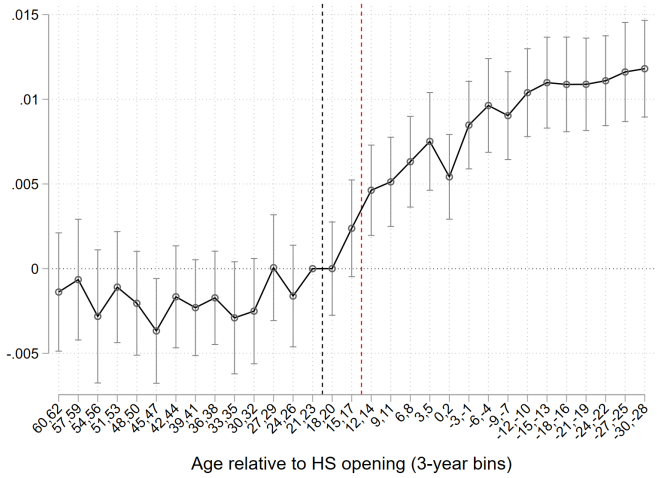
(a) Baseline (Gardner) results



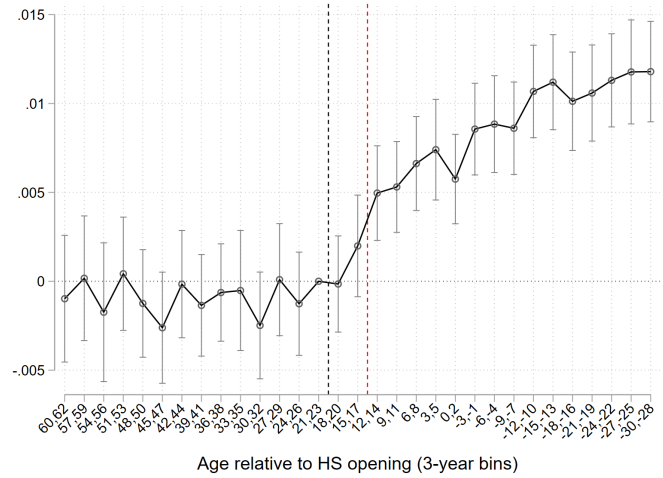
(b) TWFE estimation



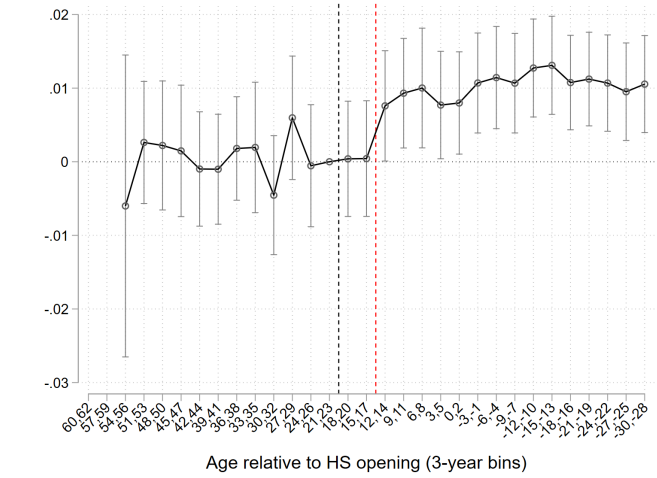
(c) Birth-year-by-state FEs



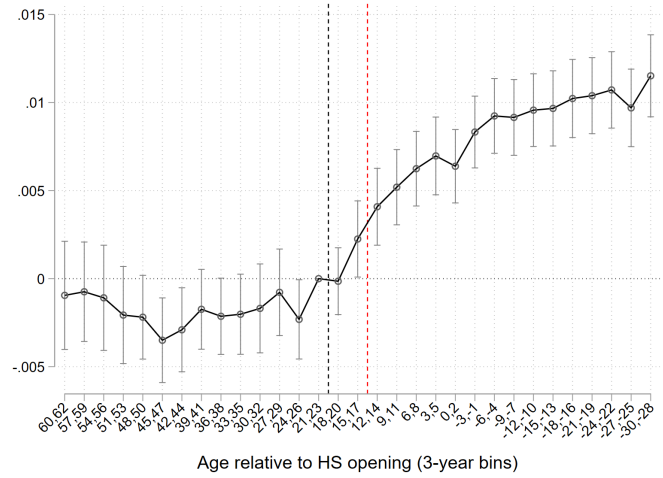
(d) Year-by-county FEs



(e) Non-movers



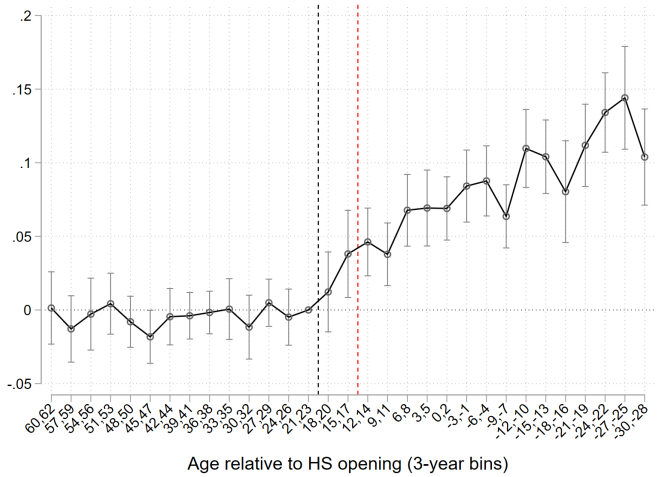
(f) 14-18 year old sample



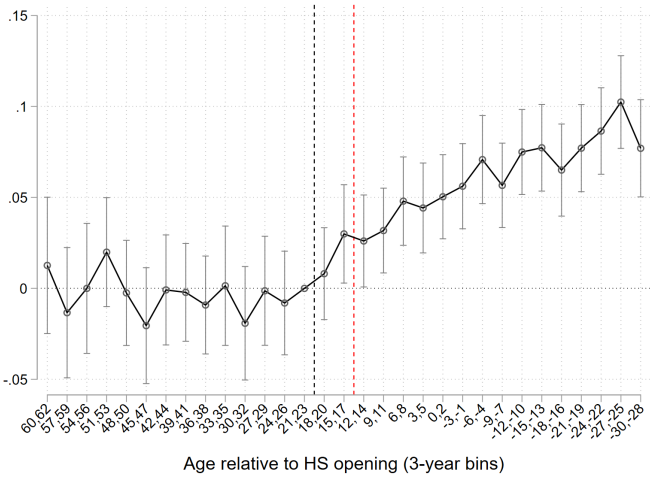
Notes: This figure shows our event study results for each indicated robustness check.

Figure A10: Logged occupational income results robustness

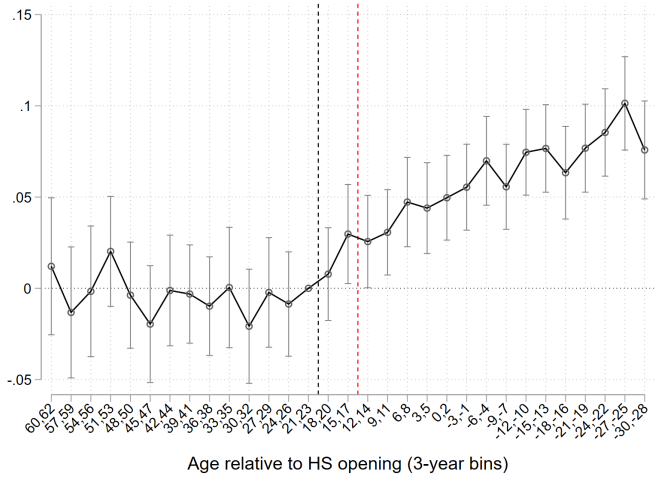
(a) Baseline (Gardner) results



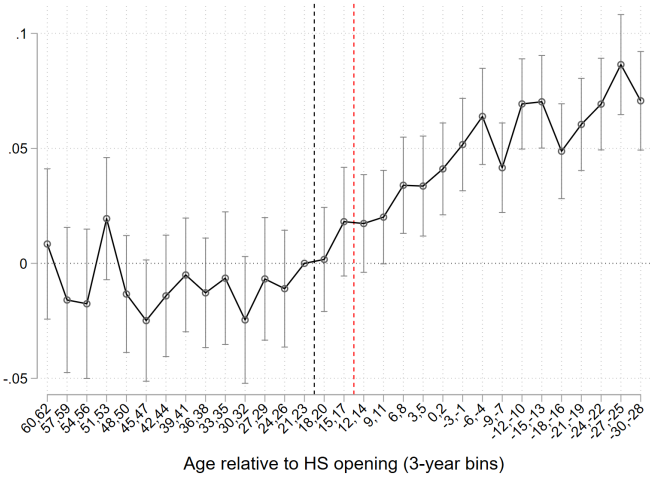
(b) TWFE estimation



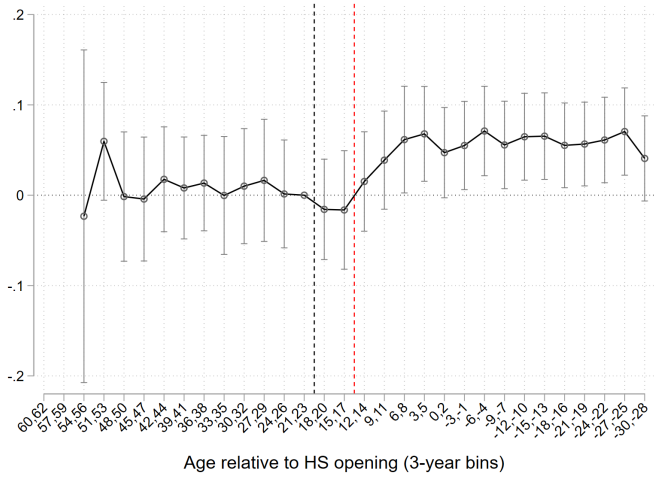
(c) Birth-year-by-state FEs



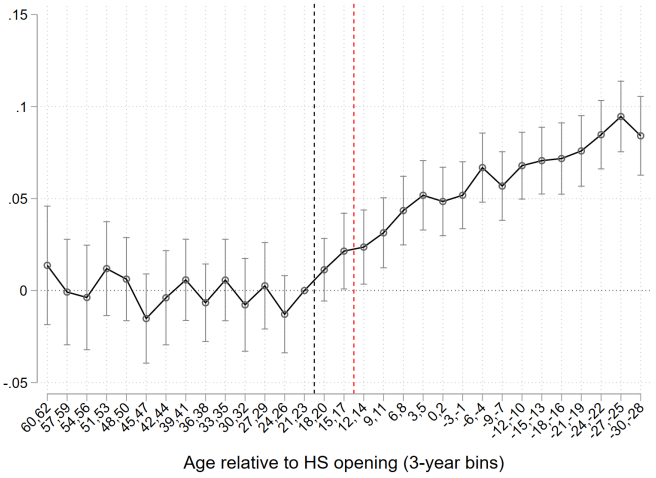
(d) Year-by-county FEs



(e) Non-movers



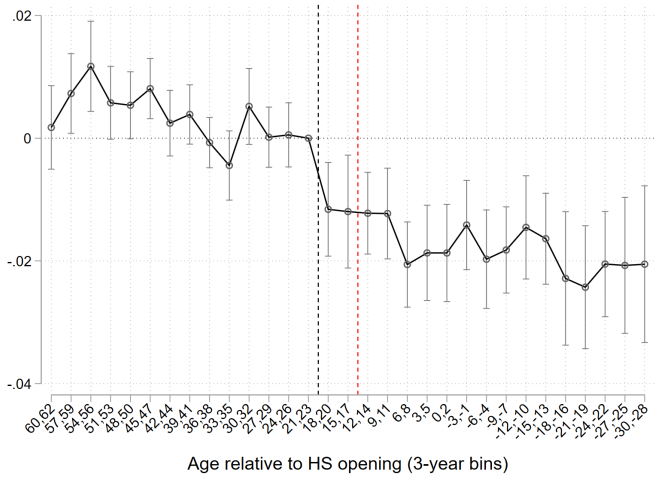
(f) 14-18 year old sample



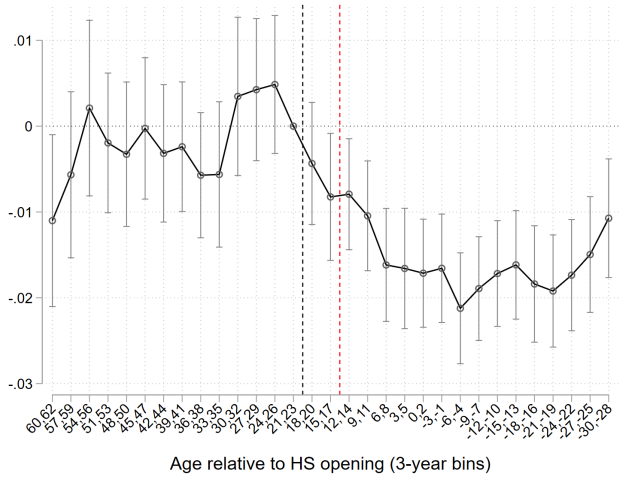
Notes: This figure shows our event study results for each indicated robustness check.

Figure A11: Marriage results robustness

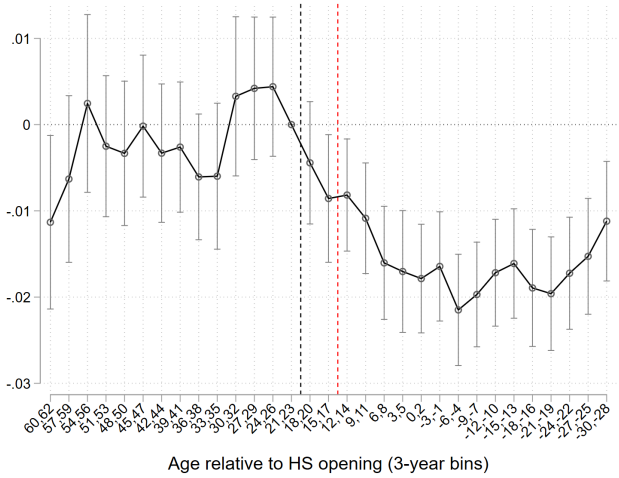
(a) Baseline (Gardner) results



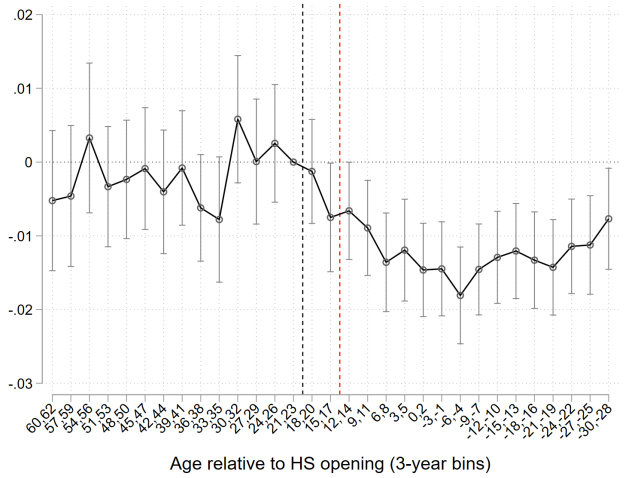
(b) TWFE estimation



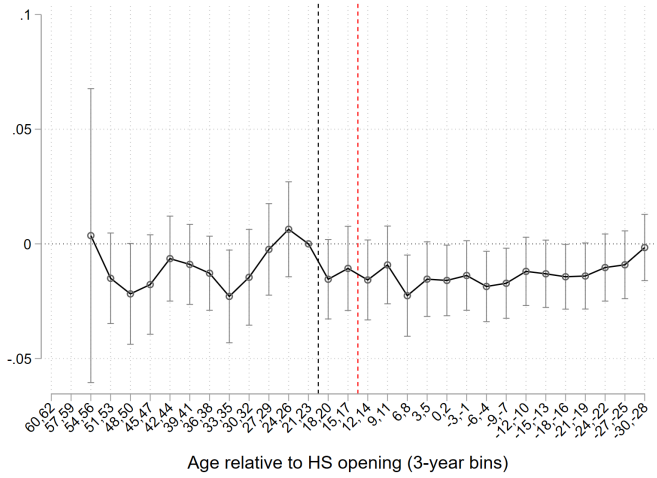
(c) Birth-year-by-state FEs



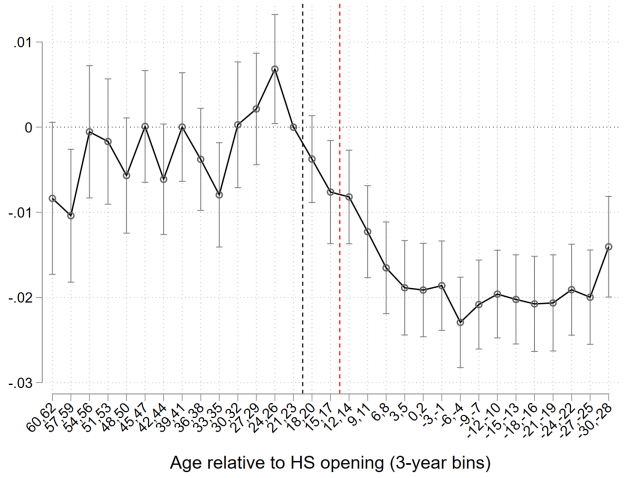
(d) Year-by-county FEs



(e) Non-movers



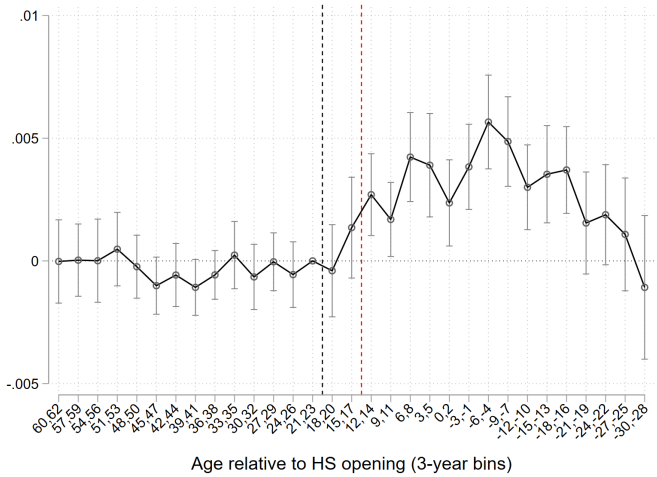
(f) 14-18 year old sample



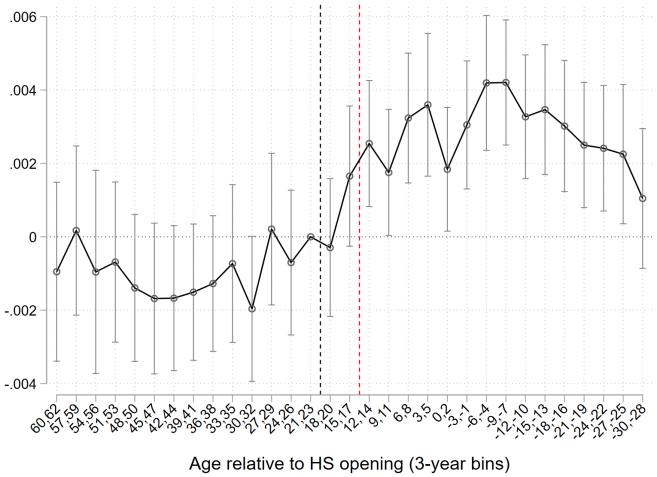
Notes: This figure shows our event study results for each indicated robustness check.

Figure A12: Teacher results robustness

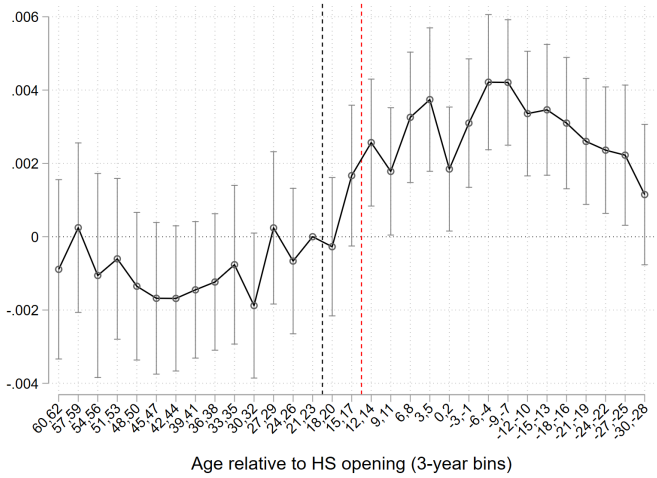
(a) Baseline (Gardner) results



(b) TWFE estimation



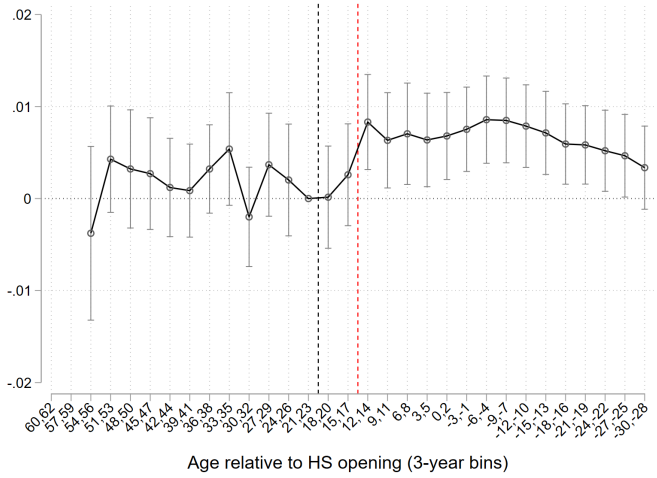
(c) Birth-year-by-state FEs



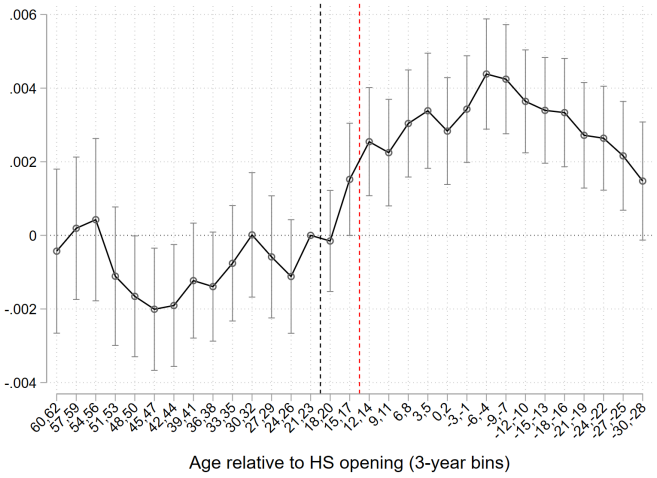
(d) Year-by-county FEs



(e) Non-movers



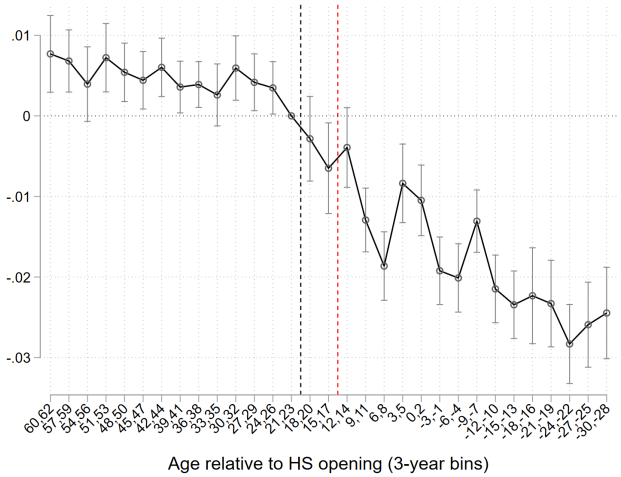
(f) 14-18 year old sample



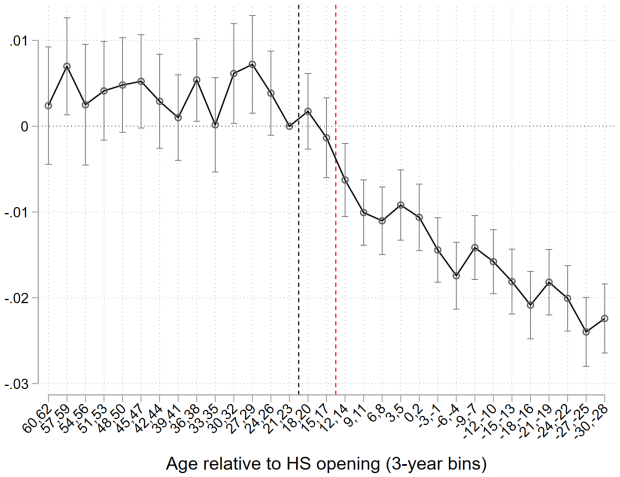
Notes: This figure shows our event study results for each indicated robustness check.

Figure A13: Farm laborer results robustness

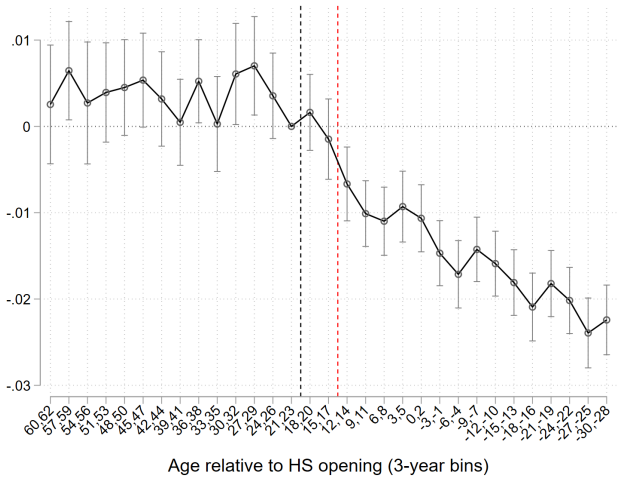
(a) Baseline (Gardner) results



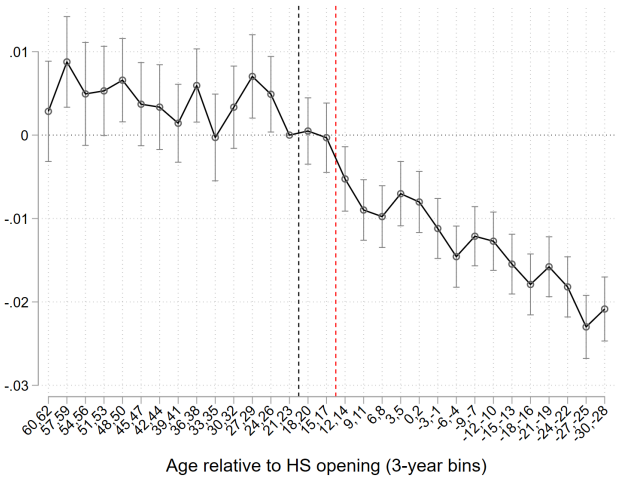
(b) TWFE estimation



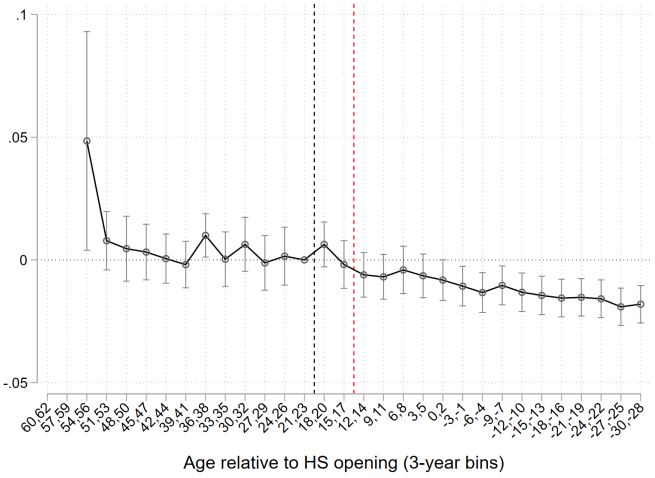
(c) Birth-year-by-state FEs



(d) Year-by-county FEs



(e) Non-movers



(f) 14-18 year old sample



Notes: This figure shows our event study results for each indicated robustness check.