Online appendix

When Sarah Meets Lawrence: The Effects of Coeducation on Women's College Major Choices

Avery Calkins, Ariel J. Binder, Dana Shaat, Brenden Timpe

A A formal model of the effect of coeducation on women's STEM majoring

We use a very simple Roy model of college and major choice to illustrate the possible effects of transition to co-education on subsequent women's outcomes. We assume there are 3 collegiate institutions in the market: h, j and k. There are two time periods: 0 and 1, which are separated by a substantial number of years. At $t = 0$, institutions h and j are women-only while k is co-educational. Between $t = 0$ and $t = 1$, institution j transitions to co-education. All institutions in each time period offer two majors: STEM (S) and non-STEM (NS). We assume away capacity constraints. (In Section F we show that evidence consistent with this assumption.)

Each time period consists of two stages. In the first stage, women make enrollment decisions η under uncertainty about the values of attending each college. In the second stage, women who have chosen to enroll in a college choose a major μ in which to graduate, with full information about major-specific payoffs. We assume that every woman enrolls in college, and that every woman who starts college completes a degree at her starting institution.

Consider a hypothetical high school senior w making decisions in period t. A given enrollment choice η_{wt} returns the expected payoff $V_{wt}(\eta_{wt})$. She chooses the enrollment choice η^*_{wt} that maximizes this function:

$$V_{wt}(\eta^*_{wt}) = \max\left\{V_{wt}(h), V_{wt}(j), V_{wt}(k)\right\}. \tag{6}$$

After making her enrollment choice, woman w realizes her major-specific payoffs and chooses her major μ_{wt}. We represent her payoff from choosing major μ at institution η as $v_{wt}(\mu_{wt}; \eta)$. Her major choice μ^*_{wt} thus satisfies:

$$v_{wt}(\mu^*_{wt}; \eta) = \max\left\{v_{wt}(S; \eta), v_{wt}(NS; \eta)\right\}, \ \eta \in \{h, j, k\}. \tag{7}$$

Woman w's expected payoff from enrolling at institution η is simply equal to the expected payoff from choosing her most-preferred major at η:

$$V_{wt}(\eta) = E\left[v_{wt}(\mu^*_{wt};\eta)\right] \tag{8}$$

Assume there are many women w in the market with varying preferences for colleges and majors. Consider the students who chose to enroll at women's institution j in period t. Denote each enrolled woman as belonging to the set A_{jt}. The share of this student body graduating from h with a STEM degree is given by $s_{STEM,jt}$:

$$s_{STEM,jt} = \frac{\sum_{w \in A_{jt}} \mathbf{1}\left\{S = \operatorname{argmax}\left\{v_{wt}(S;j), v_{wt}(NS;j)\right\}\right\}}{\sum_{w} \mathbf{1}\left\{j = \operatorname{argmax}\left\{V_{wt}(h), V_{wt}(j), V_{wt}(k)\right\}\right\}} \tag{9}$$

Suppose that, aside from institution j transitioning to co-education, nothing else changes between periods 0 and 1. Then, the object

$$\Delta = s_{STEM,j1} - s_{STEM,j0}$$

describes the treatment effect of co-education on the production of women STEM majors at institution j.

Two channels determine Δ. First, suppose that the set of women enrolling at institution j, A_j, does not change between time periods 0 and 1. Then, Δ simply depends on how the transition to co-education alters the payoffs to majoring in STEM ($v_w(S;j)$), relative to majoring in non-STEM ($v_w(N;j)$), for this population of women. We call this the "environmental effect." See Section 1.2 for a discussion of the various channels determining this effect.

Second, the transition to co-education might induce a change in the enrolled set of students A_j. To see why this might be the case, plug (8) into (6) and re-express the optimal enrollment decision:

$$\eta_{wt} = \operatorname{argmax}\left\{E\left[v_{wt}(\mu^*_{wt};h)\right], E\left[v_{wt}(\mu^*_{wt};j)\right], E\left[v_{wt}(\mu^*_{wt};k)\right]\right\} \tag{10}$$

That is, women forecast their (major-specific) payoffs from attending each institution, and use those expectations to guide their enrollment decisions.

When institution h transitions to co-education, the women that strongly desire a single-sex environment may experience a reduction in $E\left[v_{wt}(\mu^*_{wt}; j)\right]$ and may substitute from j to women's college h. Additionally, the women that strongly desire a co-educational environment may experience an improvement in $E\left[v_{wt}(\mu^*_{wt}; h)\right]$, and may substitute from co-educational college k to j. If the women who most desire a single-gender environment also have the highest expected payoffs from majoring in STEM (say, because they are the most prepared for STEM coursework), then j's transition to co-education causes its subsequent population of women to become more negatively selected on expected STEM payoffs: plausibly leading to a reduction in STEM majoring. We call this channel the "composition effect."

In Section 4, we estimate the overall treatment effect Δ. Because the assumption that nothing else about the collegiate environment changes between periods 0 and 1 is likely false, we apply difference-in-difference methodologies to estimate Δ. That is, we compare the evolution of women's major choices at colleges that transitioned to coeducation to the evolution of major choices at comparable colleges that did not transition. Section 5 attempts to decompose Δ into composition versus environmental effects.

B Data Collection and Sample Construction

B.1 Data collection on years of the switch to coeducation

Our research design requires a comprehensive timeline of the process by which historical women's colleges converted to coeducation in the latter half of the 20th century and first two decades of the 21st century. Since to the best of our knowledge there did not exist a comprehensive list of this nature, we collected the information by hand.

We define the first year of coeducation as the first year that men were admitted to traditional four-year undergraduate programs with coeducational courses. Schools where men were admitted to these programs only as commuter students are counted as coeducational, but schools where men could only participate in evening or adult education classes or graduate programs are not.

We sourced the years that single-sex institutions switched to coeducation in three different ways. The first source of information was a comprehensive check of the top 120 liberal arts colleges and the top 80 universities in the 2018 *U.S. News and World Report* for the gender of the student body in 1966 and a date of switch to coeducation. The second source of information was a list of current and former women's colleges from the Women's College Coalition, including a date of switch to coeducation. Finally, we generated a list of institutions that awarded more than 90% of their degrees to women in the first year they appeared in the HEGIS/IPEDS data and investigated these institutions by hand using a variety of resources, including Howe, Howard and Strauss (1982) and institutions' own websites. Over 90% of our transition dates were found on *.edu* websites. The three lists were then compared. Institutions that appeared on multiple lists with matching switch dates were considered confirmed. Institutions with conflicts between the switch dates or that appeared on only one list were independently verified. This procedure identified 211 institutions that were women-only in the first year they were observed, 154 of which eventually transitioned to coeducation.

We thank Claudia Goldin and Lawrence Katz for sharing a similar, independently collected dataset that covers a partially overlapping time period (the late 1800s to roughly 1990; see Goldin and Katz (2011)). The transition dates for most former women's colleges are consistent across the datasets; where they disagree, the discrepancies are usually only 1-2 years or can be attributed to differing definitions of coeducation.

B.2 Constructing our sample

We are interested in studying the effect of a rapid influx of male students into a historically female-only college campus. Our original sample consists of 154 "switching" institutions and 3,663 potential comparison schools. Many of these institutions are outside the population of interest for this paper (e.g., junior colleges, art schools) or did not offer a setting that provides a "clean" transition from single-sex to coeducation (e.g., coordinate schools that had long allowed the female-only study body to take classes at a nearby male-only college). After making a number of restrictions to narrow our sample to the population of interest, we are left with a treatment group of 77 schools and 934 comparison schools. The sample restrictions, and their impact on the eventual analysis sample, are detailed below. The restrictions' impact on our sample size applies if these restrictions are implemented in order; some schools may satisfy multiple criteria for exclusion.

1. First, we restrict the sample to institutions that were female-only or coeducational in the 1965-66 school year, the first year in which we observe degree completions. We also drop schools that had converted from male-only to coeducation in the period shortly before our sample begins. This eliminates 136 potential comparison schools. (Resulting sample includes $N_c = 3,527$ potential comparison schools and $N_t = 154$ treated schools.)

2. To ensure we observe a reasonably lengthy pre-period for our event-study estimates, we eliminate treated schools that we see for fewer than 4 years prior to the switch to coeducation or that are completely missing from the data durign this pre-period. This means the earliest transition to coeducation in our analysis sample is the 1969-1970 school year. These restrictions eliminate 24 treated schools from the sample, as well as 7 potential comparison schools that had transitioned between 1954 and the start of our sample. ($N_c = 3,520$, $N_t = 130$)

3. To allow us to observe at least a decade of post-transition outcomes, we remove women's colleges that adopted coeducation after 2007 from our treatment group. We retain these institutions as potential *comparison* schools. This restriction removes 12 colleges from our treatment group but adds them to the pool of comparison schools. ($N_c = 3,532$, $N_t = 118$)

4. To ensure our sample is limited to schools that switched from female-only to mixed-gender environments, we eliminate schools that were ever classified as coordinate institutions or merged with a men's college. We made this restriction because we suspect that classes on campus were coeducational long before mergers occurred, as is common with coordinate institutions. This restriction eliminated 23 treated schools and 27 untreated schools from the sample. ($N_c = 3,505$, $N_t = 95$)

5. We drop schools that entered the data after 1987. The IPEDS data dramatically expanded the sample at this time to include schools that had not been classified as "institutions of higher education" under Title IV of the Higher Education Act of 1965 and the response rate of those new institutions was much lower than the response rate of institutions included in HEGIS. Most of these schools were community colleges or other similar institutions. See the ICPSR documentation of the 1986-1987 academic year finance data for further details. This restriction eliminates 2 treated institutions and 1,703 untreated institutions from the sample. ($N_c = 1,802$, $N_t = 93$)

6. We eliminate for-profit institutions, once again to focus on traditional liberal arts programs. This restriction eliminates 55 untreated schools. ($N_c = 1,747$, $N_t = 93$)

7. We drop schools that closed fewer than 10 years after the switch to coeducation, as well as untreated schools that were in the data for fewer than 15 years. This restriction eliminates 6 treated and 322 untreated schools. ($N_c = 1,425$, $N_t = 87$)

8. Since our focus is on the choice of major for women at traditional liberal arts colleges, and in particular on the share majoring in quantitative fields, we eliminate schools that did not grant any degrees in STEM fields in the first year we observe them in the data. This restriction

eliminates 10 treated schools and 491 untreated schools. ($N_c = 934$, $N_t = 77$)

Returning coordinate schools and mergers (item 4), post-1987 entrants (item 5), for-profit institutions (item 6), institutions that closed shortly after the transition to coeducation (item 7), and institutions that may not have offered STEM degrees (item 8) does not substantially change our main results. We reported estimated effects on the share of women majoring in STEM using this larger sample in Appendix Figure A1. The estimated effects on the share of females majoring in STEM is very similar, although slightly smaller. This is accords with our expectations based on our reasoning for excluding these groups. The exclusion of coordinate colleges and mergers (item 4) is particularly important, as it is not clear that there was truly a transition from women-only courses to coeducational courses at either time. Especially at institutions where there was a merger between a women's college and either a men's college or an institution which was already coeducational, we think it is likely that a number of other changes came about at the same time, and coordinate colleges likely had coeducational courses before the transition to coeducation, muting the effects of the transition to coeducation. Adding schools that did not appear to have STEM programs (item 8) would also be expected to attenuate our estimates, since changes in the STEM share of female degrees would be 0 or positive by construction.

C Implementation of the Callaway and Sant'Anna (2021) estimator

Our research design exploits variation in the timing of women's colleges' switch to coeducation, as well as variation in the decision to switch at any time, to study the effect of the gender mix of the collegiate environment on women's choice of major. Because we expect the effect of this reform to evolve dynamically, we present event-study estimates that show the evolution of changes in choice of major at switching colleges relative to the comparison group.

The conventional event study model is based on a two-way fixed effects (TWFE) design, which implements a pooled panel regression with controls for unit (j) and time (t) fixed effects to estimate the impact of a policy for each time period relative to the date t_j^* of implementation of the reform:

$$y_{jt} = \alpha_j + \theta_t + \sum_{k=m}^{M} \beta_k \mathbb{1}\left\{t - t_j^* = k\right\} + \epsilon_{jt} \tag{11}$$

Recent studies have revealed that the TWFE specification may provide misleading estimates of treatment effects when there is variation in treatment timing across units, as there is in our setting (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; de Chaisemartin and D'Haultfœuille, 2020; Borusyak, Jaravel and Spiess, 2021; Goodman-Bacon, 2021). These issues are particularly pronounced in the presence of un-modeled heterogeneous effects across units.

We instead adopt a slightly modified version of the estimator proposed by Callaway and Sant'Anna (2021).[27] The estimator avoids the shortcomings of TWFE models by estimating event-study-style treatment effect parameters separately for each treatment "cohort" g (e.g., schools that switched in 1969-1970 school year are part of cohort $g = 1969$), then aggregating those cohort-specific effects into an overall estimate of the average treatment effect on the treated. By estimating the effects one cohort at a time, the procedure facilitates transparency in the choice of comparison group used for each treatment group (e.g., the researcher can ensure the comparison group is not polluted by a recently treated unit that may still be adjusting to the reform) and allows potentially heterogeneous effects to be aggregated using the choice of weights best suited for estimating target parameter of interest (i.e., it avoids weighting by the inverse of the variance of exposure to the treatment, as is the default in regression-based methods).

The doubly-robust estimator developed by Callaway and Sant'Anna (2021) relies on two strategies to construct appropriate counterfactual trends for units

[27] We are indebted to Brantly Callaway and Pedro Sant'Anna for their generous and illuminating correspondence about the finer details of their estimator.

that adopted the treatment of interest. First, to limit the comparison group to schools that "look like" the treated group, each school j gets a cohort-g-specific propensity score $\hat{p}_g(X_j)$. In addition, the counterfactual trend in outcome y_{jt} between time t and some base period b is estimated by regressing changes $\Delta y_{jt,b} = y_{jt} - y_{jb}$ on the same vector of covariates X_j in a sample made up solely of the comparison group, and then using these regression estimates to predict changes $\Delta \hat{y}_{jt,b}(X_j)$ for the treated cohort g.

Formally, for a sample made up of schools $j \in \{1, 2, ..., J\}$, the estimator is constructed with the following sample analog of Callaway and Sant'Anna (2021) equation 2.4 (weights are omitted here for parsimony, but it is straightforward to add them):

$$\hat{\alpha}_{g,t} = \frac{1}{J} \sum_{j \in J} \left(\frac{G_{jg}}{\frac{1}{J} \sum_j G_{jg}} - \frac{\frac{\hat{p}_g(X_j) C_{jg}}{1-\hat{p}_g(X_j)}}{\frac{1}{J} \sum_j \frac{\hat{p}_g(X_j) C_{jg}}{1-\hat{p}_g(X_j)}} \right) (\Delta y_{jt,b} - \Delta \hat{y}_{jt,b}(X_j)) \tag{12}$$

where G_{jg} is a binary indicator for school j belonging to treatment cohort g and C_{jg} is an indicator for belonging to the pool of candidate comparison schools for group g.

Our implementation differs in a few minor ways from the procedure outlined by Callaway and Sant'Anna (2021). The first is the choice of base period b. The estimator used by Callaway and Sant'Anna (2021) sets the base period as the year before researchers believe treatment effects would be expected to arise in their setting. In cases without anticipation effects, this would generally mean $b = g - 1$. We instead define y_{jb} as the average of the outcome variable in the five years immediately preceding the switch to coeducation, $y_{jb} = \frac{1}{5} \sum_{s=g-5}^{g-1} y_{js}$. This choice requires slightly stronger assumptions about parallel trends (Marcus and Sant'Anna, 2021), but should improve efficiency and reduce the impact of noise on our estimates (Borusyak, Jaravel and Spiess, 2021).

Second, the conventional estimator – as implemented in the "did" package for R – defines $\Delta y_{t,b} = y_t - y_b$ for $t \geq t^*_j$, but as single-year differences $\Delta y_{t,b} =$

$y_t - y_{t-1}$ for $t < t_j^*$. We instead adopt the former definition for all periods, i.e., $\Delta y_{t,b} = y_t - y_b \; \forall \; t$. While the two approaches are very similar conceptually, we believe our approach to reporting event-study results will be more familiar and intuitive for our readers, most of whom are accustomed to interpreting event-study coefficients as changes in an outcome relative to an omitted period, rather than as the first derivative of those changes.

Finally, rather than estimating a propensity score with a logit or probit model for each cohort g, we use discrete variables (or discretized versions of continuous variables) to find exact matches, school by school, for each institution in the treatment group before aggregating our estimates. This is equivalent to defining a propensity score using fully saturated OLS. This approach avoids the pitfalls of estimating logit and probit models in situations with few treated units (Albert and Anderson, 1984; Firth, 1993). It also allows us to focus on what we believe are the most important conditioning variables in our setting, which happen to be discrete.

These three changes result in a simplified version of equation 12. To see this, note that because our vector X_j is made up entirely of discrete variables, the propensity score for any group g will either 0 or a constant $\bar{p}_g$, where $\bar{p}_g$ is the share of treated schools among all institutions where X_j is identical to the treated school in question. In addition, because we define group g as a single school, the formula simplifies to

$$\hat{\alpha}_{g,t} = (\Delta y_{jt,b} - \Delta \hat{y}_{jt,b}(X_j)) - \sum_{j \in J} \omega_j(X_j) \left(\Delta y_{jt,b} - \Delta \hat{y}_{jt,b}(X_j) \right) \tag{13}$$

where $\omega_j(X_j)$ sums to 1 and represents school j's share of the sample for which $C_{jg} = 1$ *and* $p_g(X_j) > 0$, i.e., its share of the comparison group for treatment group g. Since $\Delta \hat{y}_{jt,b}(X_j)$ is calculated using only candidate comparison schools with strictly positive propensity scores, two of the final three terms cancel out and equation 13 simplifies to equation 2.

In our preferred specification, our comparison group consists of all women's colleges that switch to coeducation at least 10 years after cohort g – or never switch at all. Our vector X_j includes indicators for affiliation with the Catholic church and having a selectivity ranking of 1-3 in the 1972 Barron's ratings. Our estimates rely on the assumption that, conditional on these observed characteristics, trends in women's choice of major at our comparison group of schools accurately reflects the counterfactual trends that would have occurred at our treated schools in the absence of a switch to coeducation. As robustness checks, we also estimate $\alpha_{g,t}$ using only never-treated women's colleges as the comparison group, and then by using *all* four-year colleges that did not switch the gender mix of their student body during our sample period. Because the latter exercise adds a large number of schools to the comparison group – many of which are very different from our treated group of historical women's colleges – we add two additional characteristics to our vector of covariates X_j: A measure of school size (proxied by discrete categories of the number of degrees awarded in pre-reform years) and pre-reform linear trends in the share of degrees among all students that are in STEM fields. All estimates are weighted by the number of female students in the school in its first year in our dataset.

Callaway and Sant'Anna (2021) also propose an inference procedure that accounts for multiple hypothesis testing across time periods in an event-study figure. While the results in our main appendix report pointwise 95% confidence intervals, Figures A6 and A7 show that our estimates are noisier but, for the most part, still statistically distinguishable from 0 when using this procedure. Multiple-testing corrections have minimal impact on our estimates of "long-run" effects.

D Robustness check: the synthetic control method

As a robustness check on our main result, we use the synthetic control method to estimate the effects of transitioning to coeducation on women's STEM major choices. The synthetic control method offers a data-driven procedure to construct a control group that matches our treatment group based on pre-treatment characteristics. Thus, it may provide a valid comparison group even if our identification assumption fails in the standard difference-in-differences methodology used above.

One complication of our setting is that we have multiple "treated" schools rather than the single treated unit that is standard in synthetic control settings (e.g. Abadie, Diamond and Hainmueller, 2010). We adjust the standard procedure in two ways to incorporate this complication. First, we group schools that switched to coeducation in the same year, so that the "treated" groups are effectively school-cohort combinations. Second, we construct a synthetic control group separately for each cohort of treated schools and then average the effects by year relative to the switch (Cavallo et al., 2013; Acemoglu et al., 2016).

Our baseline specification constructs a synthetic control group for each treated school-cohort observation by matching on the entire set of pre-*transition* outcome variables (Ferman, Pinto and Possebom, 2020). Appendix Figure A8 reports the results of this estimation procedure. For consistency with our event-study results, time 0 corresponds to the effect on the female STEM share in the junior year the first coeducational cohort. Note that because years -2 and -1 are not used in the matching procedure, the fact that they remain near 0 provides an informal cross-validation test and some reassurance of the validity of our design. In fact, we see little evidence of a departure from 0 effect until the graduating year of the first coeducational senior class. The synthetic control event study traces a similar path as did our standard event study (Figure A2): it shows a 2 percentage point decrease in the share of women majoring in STEM by five years after the transition to coeducation and a 3 percentage point decrease by nine years after the transition to coeducation. We calculate a "difference-in-differences" estimate by averaging the post-treatment coefficients and subtracting them from the average pre-treatment coefficients. The estimate of -0.025 is an outlier in the distribution of placebo effects, with a p-value of 0.01.[28] This estimate is slightly larger in magnitude than the one

[28] We conduct inference by randomly reassigning treatment status and estimating the effect of the transition to coeducation on the placebo institutions, using 250 replications (Abadie, Diamond and Hainmueller, 2015). If our estimated effect is either below the 2.5th percentile or above the 97.5th percentile of placebo effects, the effect is statistically significant.

we obtain in our main event study model.

E Other data processing notes

E.1 Major codes

E.1.1 Coding scheme and crosswalks

This paper uses consistent 4-digit, 2-digit, and grouped 2-digit versions of major codes. The consistent coding scheme is based on the 1990 version of the Classification of Instructional Programs (CIP) from the National Center from Education Statistics (NCES).

Codes to describe college majors have been revised several times over our sample period. There were two sets of major codes in the HEGIS data, with a revision in 1970, and coding switched to the CIP in the early 1980s.[29] Revisions of the CIP occurred in 1985, 1990, 2000, and 2010.[30] Crosswalks between the 1970s HEGIS codes and the CIP, and between different versions of the CIP, are available from NCES, but they are not complete.

Similar to occupation codes, the CIP has 2-, 4-, and 6-digit versions of codes, while the HEGIS codes have only 2- and 4-digit versions. Revisions of the CIP only rarely move major categories across 2-digit codes,[31] though the 1990, 2000, 2010 revisions did move, split, and combine some two-digit codes.[32]

For this paper, all 6-digit codes were crosswalked to the 4-digit 1990 CIP. Where crosswalks provided by the NCES were incomplete, they were supple-

[29]The first version of the CIP was constructed in 1980, but HEGIS seems not to have adopted it until 1983.

[30]There seems to have been late adoption of the new coding schemes in the IPEDS data – the switches seem to have occurred in 1987, 1992, 2002, and 2012, and may not have occurred uniformly across schools. Revisions of the CIP vary in how many changes were made, with the 1985 revision being much smaller than subsequent revisions.

[31]Exceptions include clinical versions of the life sciences, materials science, and educational psychology, all of which could be considered to be part of multiple two-digit codes.

[32]For instance, the 1990 revision of the CIP combined category 17, Allied Health, with category 18, Health Sciences, into category 51, Health Professions and Related Sciences. Most of the 4-digit categories were preserved but re-numbered in the revision.

mented by lists and descriptions of CIP codes created by the NCES. When majors were not included in the NCES crosswalks, they were matched to the major of the most similar title and description in the 1990 CIP. If two 4-digit codes were combined in any version of major codings after 1970, they were combined in the consistent coding scheme. The same is true for the 2-digit codes. Six-digit majors that were created or deleted at any point were assigned to the same 4-digit code in the "other" category, and 4-digit codes that were ever created or deleted were assigned the the 4-digit code for "other" within the same 2-digit code.[33] Four-digit majors with fewer than 950 school-by-year observations were combined with majors that cover similar material[34] or with the "other" category within their two-digit code. Smaller 2-digit codes, such as Law, Library Science, and Military Science, were treated as a single 4-digit code.

For the main result, majors were combined into groups of 2-digit codes, with the most important of those groups being STEM. STEM in this case includes the 2-digit codes for Life Sciences, Physical Sciences, Engineering, Computer Science, and Mathematics. Alternative specifications also included Health Professions.

E.1.2 Categories of majors

The following list is the two-digit categories of majors in each group of 2-digit codes. Groups are in bold and the two-digit categories are listed afterward. Where the two-digit sets of codes are not informative, four-digit codes are included in parentheses. Some groups contain only one two-digit code. The "other" group includes majors that generally cannot be found at small liberal arts colleges or that are generally very small.

Art Visual and performing arts, architecture and related services

[33] For instance, African Languages were not included in the 1990 CIP and were therefore assigned to the 4-digit code for Other Foreign Languages.

[34] For instance, Architectural Engineering and Civil Engineering, Business Administration and Enterprise Management, and the health categories such as medicine, dentistry, and others which require a professional degree.

Business Business, marketing

Education All education fields (including math education)

Economics Economics (4-digit code)

Health Health professions and clinical services

Home Economics Home economics/family and consumer sciences

Humanities Area and group studies (e.g. gender studies, Hispanic Studies), English, foreign languages and linguistics, philosophy and religious studies

Psychology Psychology

Other Social Sciences Social sciences except economics (general social science, anthropology, criminology, demography, geography, history, international relations, political science, social science, urban studies), communications

STEM Life sciences, physical sciences, mathematics and statistics, computer and information science, engineering, engineering technology, science technology

Other Agriculture, forestry, law, trades/vocational, military science, library science, multi- and inter-disciplinary, theology and religious vocations, protective services, public administration and social services

E.2 School Codes

NCES uses two different coding schemes for individual schools at different points in the data. HEGIS identifies schools using FICE codes, which is a six-digit identification code assigned to schools doing business with the Office of Education in the 1960s. IPEDS uses the UnitID, which is also a six-digit code. Our data uses the FICE as a consistent identifier throughout the survey, with some modifications as detailed below.

Not every institution has a FICE code. Institutions that do not have a FICE code are those that entered the IPEDS data after the Institutional Characteristics file stopped listing FICE codes (which was during the 1990s). We drop those institutions from our sample, as according to the ICPSR files for IPEDS financial characteristics between 1988 and 1990, institutions that entered the sample after the beginning of the IPEDS have a much lower response rate than institutions in the HEGIS sample. However, the data set itself has the UnitID entered in place of the FICE code for those institutions.

Some institutions have multiple FICE codes. In most of these cases, a public institution originally reported all branches under one observation, and then switched to reporting each branch separately. The vast majority of cases where all degrees awarded are reported under the main campus occur in 1966, with a few additional cases between 1967 and 1969. We do not link such cases together. In other cases, an institution switched FICE codes in the middle of the sample. We are generally not sure why this occurs. We do link these cases together so that we have a single FICE code for all years the institution was in the data. Finally, there are a few institutions (notably Cornell and Columbia) with several different administrative units that separately report degrees awarded to IPEDS and HEGIS. We treat these institutions as a single observation and collapse them to a single FICE code.

Some FICE codes apply to multiple institutions. In these cases, all institutions are part of the same system, and the majority of these cases occur among institutions who enter the data in 1987 and later, especially among for-profit institutions with multiple campuses nationwide (e.g. the University of Phoenix). There are some cases where a public college with several branches (e.g. the University of Pittsburgh) reported degrees separately from each branch but reported the same FICE from each school. Where we could, we assigned these institutions to separate codes for each branch, but the rest of them are collapsed to the FICE level. We have also dropped schools that are ever classified as for-profit schools from our sample, which removes many of these cases from our analysis.

F The role of capacity constraints

One possible explanation for our finding of a reduction in the share of women majoring in STEM at newly coeducational colleges is that such colleges hit capacity constraints to a larger extent in STEM than in non-STEM fields. For example, if STEM is both more costly for colleges to provide and more popular among men, students might be more crowded out of STEM majors than non-STEM majors after the transition to coeducation. This would suggest a negative relationship between field-specific cost of instruction and growth in degrees earned.

To test this hypothesis, we examine growth in degrees earned (by both men and women) in a wide range of fields and compare these figures with the marginal cost of instruction, as reported by Hemelt et al. (2018). In the first step, we use equation 5 to estimate the long-run effect on the number of degrees awarded in each field. We then link these results to estimates of marginal costs. Where the fields in our data did not overlap exactly with those reported in Hemelt et al. (2018), we aggregate the marginal cost estimates by calculating the simple average.[35]

Results are shown in Figure A11. The four main fields of interest in our paper are shown as orange triangles, while all others are in blue. Among the three STEM fields of math, biology, and physical sciences, we see a slight *positive* relationship, suggesting that the costliest-to-teach fields were also the fields where growth was largest. Indeed, despite the relatively low cost of adding students to math or economics courses, we find significant negative effects on the share of women majoring in these fields, suggesting that these effects cannot be explained by women being physically crowded out of the classroom.

[35] Hemelt et al. (2018) report a range of cost estimates by field. We rely on marginal costs estimated with program fixed effects, and using only schools without graduate programs to maximize comparability with the treated schools in our sample. For their estimates, see Table 5, column 5 of their NBER working paper.

G The National Longitudinal Study of the Class of 1972

We conduct additional analyses using the National Longitudinal Study of the Class of 1972 (NLS72) to determine which baseline (i.e., pre-freshman) characteristics are most important for predicting that a student will complete a STEM degree (National Center for Education Statistics, 1999). We chose the NLS72 for this analysis because it provides information on both characteristics before the beginning of college and STEM degree completion for a cohort who attended college during the part of our sample period when most of our switching colleges transitioned. NLS72 is a nationally representative longitudinal survey that followed high school seniors in the graduating class of 1972 for twelve years following completion of high school. The baseline survey, conducted in spring 1972 (right before students graduated high school), collected a substantial amount of information on students' backgrounds and plans for the future. The five follow-up surveys (conducted between 1973 and 1984) focused on what students had been doing since the previous survey (including degree completion and college major). We focus on female students who responded to both the baseline survey and the fourth follow-up (conducted between October 1979 and May 1980), who indicated on the baseline survey that they planned to attend a four-year college starting in Fall 1972 and who had completed a degree by October 1979.

We used the NLS72 data to estimate the correlation between STEM degree completion and baseline intention to major in a STEM field, defining STEM to include biology, computer science, engineering, mathematics, and physical science (similarly to our analyses of IPEDS and HERI data). We estimate the following equation using OLS:

$$STEMBA_i = \alpha STEMIntent_i + X_i\beta + \varepsilon_i \tag{14}$$

where $STEMBA_i$ is a 0/1 indicator that the student's completed bachelor's degree was in a STEM field and $STEMIntent_i$ is an indicator variable that the student planned to major in STEM at baseline, and X_i is a vector of control variables. In our first specification, X_i includes only a constant. In our second specification, we add a set of controls for students' occupation plans (professional, homemaker, or other) and whether student considered marriage and family to be "very important." In our third specification, we add a set of controls for students' background, including an indicator that students are white, an indicator for being a first-generation college student (i.e., neither parent attended college), an indicator that the student's father completed college, an indicator that the student's mother completed college, and information on parents' occupations (indicators for a professional occupation fathers and indicators for professional occupation or homemaker for mothers). In our fourth specification, we add an indicator that students had a GPA of A- or better in high school and the number of years of high school math and science that the student completed.[36] In our subsequent analysis, we use the coefficient estimates from these regressions to predict the share of entering freshman women in our HERI Freshman Survey data who will major in STEM, and then evaluate the effect of coeducation on this predicted share.[37] We interpret this analysis as a test of composition effects.

We use the R^2 values from these regressions to determine the relative importance of each baseline characteristic to completing a STEM degree. See Panel A of Table 4 for the results. Plans to major in STEM right before high school graduation have a R^2 value of 0.191, indicating that baseline preferences account for approximately 19% of variation in STEM degree completion in the NLS72 sample. Adding each successive set of control variable shifts the R^2 by no more than 0.012, with the largest shift coming from the addition of high school grades and coursework.[38] We take these results as confirmation

[36] We recoded the control variables in NLS72 to match information available in TFS as closely as possible.

[37] In our HERI data, school-year observations sometimes have missing values for some of the characteristics measured in the NLS72 analysis. In those cases, we use all available characteristics to predict the share of women who will major in STEM.

[38] Further analyses suggest that coursework is more important than grades, consistent with Card and Payne (2017). However, information on high school coursework is only available in TFS after 1984.

that preferences are the main characteristic of interest in determining whether shifts in the composition of female students could be responsible for the effects of coeducation on future STEM majoring.

H Alternative approach to quantifying the environmental effect

Our main analysis in section 5.4 relies on estimates from our linked HERI data and the NLS72 to quantify the potential role played by composition effects in our main findings that coeducation reduced the share of women majoring in STEM. In this section, we present results from an alternative approach that relies on the assumption that women who matriculated *before* coeducation was adopted may comprise a sample that is relatively free from selection into the college. In particular, the sophomore class during the first year of coeducation would have experienced a relatively pure "environmental" effect – since, as underclasswomen, they would likely share classrooms and social spaces with the new men – but had already chosen their college. In contrast, freshman cohorts may have been more prone to composition effects, while juniors and seniors would have been both less likely to interact with the entering men and more likely to be locked into an academic program by the time those men arrived.

Our IPEDS data does not provide information on the time to degree, limiting our ability to measure cohorts precisely. Our best proxy is to focus on women who graduated in year $t_j^* + 2$, i.e., the third year of coeducation at school j. Assuming four years to completion of the degree for most women, this cohort should be made up primarily of women who were sophomores when coeducation was implemented.

Table A8 reports estimates of $\beta_{\tau=2}$ and β_{LR}, constructed from equation 3 with the share of women majoring in STEM as the outcome. Column 1 presents estimates of the effect on the sophomore class. These estimates are generally less precise than our main estimates of the long-run effect, but they can be statistically distinguished from 0 in every comparison group.

What does this imply for the question of whether the main effects are driven by composition or features of the campus environment? Note that the smaller magnitudes in column 1 of Table A8 are consistent with our conclusions that the decreasing share of women in STEM is most likely driven by increasing interactions with men, because the male share of men at former women's colleges also increased gradually. The sophomore class was thus exposed to a lower "dose" of men. If we re-scale our estimates by the male share of graduates, we find that that share of women in STEM falls by about 1.8 percentage points for every 10-percentage-point increase in men (column 1 of panel A). This figure is very similar to the estimate of 1.6 percentage points (column 2, panel A). These very similar magnitudes are far from conclusive, but they provide yet more evidence that is consistent with the hypothesis that composition effects were negligible – and, if anything, wrong-signed – in this setting.

Appendix references

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105(490): 493–505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method." *American Journal of Political Science*, 59(2): 495–510.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton. 2016. "The value of connections in turbulent times: Evidence from the United States." *Journal of Financial Economics*, 121: 368–391.

Albert, A, and JA Anderson. 1984. "On the existence of maximum likelihood estimates in logistic regression models." *Biometrika*, 71(1): 1–10.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic natural disasters and economic growth." *Review of Economics and Statistics*, 95(5): 1549–1561.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2020. "Cherry picking with synthetic control." *Journal of Policy Analysis and Management*, 39(2): 510–532.

Firth, David. 1993. "Bias reduction of maximum likelihood estimates." *Biometrika*, 80(1): 27–38.

Goldin, Claudia, and Lawrence F. Katz. 2011. "Putting the "Co" in education: Timing, reasons, and consequences of college coeducation from 1835 to the present." *Journal of Human Capital*, 5(4): 377–417.

Howe, Florence, Suzanne Howard, and Mary Jo Boehm Strauss. 1982. *Everywoman's Guide to Colleges and Universities.* The Feminist Press.

Marcus, Michelle, and Pedro H. C. Sant'Anna. 2021. "The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics." *Journal of the Association of Environmental and Resource Economists*, 8(2): 235–275.

National Center for Education Statistics. 1999. "National Longitudinal Study of the Class of 1972." Inter-University Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/ICPSR08085.v1.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2020. "IPUMS USA: Version 10.0 [dataset]." IPUMS, Minneapolis, MN.

I Tables and Figures

Table A1: Summary statistics, treated and alternative comparison schools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Treated	Candidate comparison groups				Matched comparison groups			
		Never-treated wom. clg.	Difference (1)-(2)	Liberal arts colleges	Difference (1)-(4)	Never-treated wom. clg.	Difference (1)-(6)	Liberal arts colleges	Difference (1)-(8)
STEM share of women's degrees	0.10	0.11	-0.01	0.11	-0.00	0.09	0.01	0.11	-0.00
	(0.05)	(0.05)	[0.27]	(0.06)	[0.79]	(0.05)	[0.31]	(0.06)	[0.77]
Annual growth rate, STEM	-0.005	-0.003	-0.003*	-0.003	-0.003**	-0.004	-0.001	-0.005	-0.000
	(0.011)	(0.009)	[0.07]	(0.014)	[0.04]	(0.009)	[0.53]	(0.013)	[1.00]
Total enrollment	1226	1250	-24	1500	-274**	1254	-28	1422	-196
	(917)	(653)	[0.88]	(1177)	[0.03]	(585)	[0.85]	(773)	[0.20]
Female share of all degrees	0.97	0.99	-0.02**	0.54	0.44***	0.99	-0.02*	0.60	0.37***
	(0.07)	(0.02)	[0.01]	(0.17)	[0.00]	(0.03)	[0.07]	(0.23)	[0.00]
Graduate degrees awarded	27	19	9	10	17***	14	13*	14	13
	(51)	(42)	[0.30]	(40)	[0.00]	(38)	[0.08]	(56)	[0.11]
Private college	0.92	1.00	-0.08**	0.89	0.03	1.00	-0.08**	0.97	-0.05
	(0.27)	(0.00)	[0.01]	(0.32)	[0.38]	(0.00)	[0.01]	(0.17)	[0.10]
Ever Catholic-affiliated	0.64	0.30	0.33***	0.06	0.58***	0.64	0.00	0.64	0.00
	(0.48)	(0.46)	[0.00]	(0.23)	[0.00]	(0.48)	[1.00]	(0.48)	[1.00]
Selective admission	0.19	0.45	-0.27**	0.20	-0.02	0.19	0.00	0.19	0.00
	(0.39)	(0.50)	[0.01]	(0.40)	[0.74]	(0.39)	[1.00]	(0.39)	[1.00]
Observations	77	29	106	362	439	29	106	350	427

Notes: Column 1 shows means (standard deviations) for our main treated sample of women's colleges, averaged over the five years prior to each treated school j's transition to coeducation. Comparison group means in column 2 are constructed by matching each treated school j with all women's colleges that never transitioned. We then stack these comparison groups across all treated schools j and compute sample means for the resultant "grand" group. Column 3 reports result of test of difference in means between treated (column 1) and never-treated comparison group (column 2). Column 4 reports summary statistics for comparison group of untreated liberal arts college using Carnegie Classification from 1987. Column 5 presents results of test of difference in means between treated group and untreated liberal arts colleges. Columns 6-9 report analogous summary statistics using matched samples as described in Section 3.1. STEM share of women's degrees is weighted by total degrees awarded to females, female share of degrees is weighted by total degrees awarded, and all other means are unweighted. Data are drawn from IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. See Section 2 for further information on sample construction and characteristics.

Table A2: Summary statistics for schools that switched to coeducation

	(1) All switchers	(2) Excluded switchers	(3) Main sample	(4) HERI subsample
STEM share of women's degrees	0.096	0.070***	0.103*	0.106
	(0.053)	(0.050)	(0.051)	(0.055)
Annual growth rate, STEM	-0.004	-0.002***	-0.005**	-0.003***
	(0.010)	(0.008)	(0.011)	(0.007)
Total enrollment	1190	1136	1226	1072
	(1069)	(1265)	(917)	(484)
Female share of all degrees	0.96	0.93	0.97	0.98
	(0.08)	(0.10)	(0.07)	(0.04)
Graduate degrees awarded	49	89	27	26
	(251)	(419)	(51)	(46)
Private college	0.95	1.00	0.92	0.97
	(0.23)	(0.00)	(0.27)	(0.18)
Ever Catholic-affiliated	0.56	0.40	0.64	0.63
	(0.50)	(0.49)	(0.48)	(0.48)
Selective admission	0.14	0.05	0.19	0.23
	(0.34)	(0.21)	(0.39)	(0.42)
Institutions	118	41	77	30

Notes: Table shows sample means (standard deviations) calculated in the five years prior to the switch to coeducation. Column 1 includes all women's colleges that adopted coeducation after 1965-66 and that we observe during the five years before and decade after the transition. Column 2 includes schools we drop because they did not fit into our target population of institutions that offered an arts-and-sciences curriculum and experienced sharp transitions to coeducation. Column 3 shows our main sample of colleges that switched between 1969 and 2007. Column 4 show summary statistics for the subset of column 3 that can be linked to schools in the HERI Freshman Survey. In columns 2-4, the designation of 1, 2, or 3 stars indicates that a test of differences between each subset of switchers and the sample of all switchers (column 1) results in a p-value below 0.10, 0.05, or 0.01, respectively. Data drawn from IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. Trends in STEM and total degrees are the estimated linear trend in the five years prior to the switch to coeducation. Catholic affiliation is coded as 1 if the school was ever affiliated with the Catholic Church. Schools are coded as having selective admission if they received a Barron's rating of 1, 2, or 3 in 1972. The majors included in the STEM field are described in Section 2 and Appendix B. See Section 2 for further detail on sample construction.

Table A3: Long-run effect of coeducation on presence of men at former women's colleges

	(1) Male share of freshmen	(2) Male share of students	(3) Male share of degrees	(4) Male share of faculty
Panel A: Not-yet-treated comparison group				
Long-run effect	0.211***	0.216***	0.192***	0.045***
	(0.007)	(0.016)	(0.016)	(0.005)
Counterfactual mean	0.030	0.034	0.036	0.385
Observations	5,158	5,164	5,505	2,428
Panel B: All-college comparison group				
Long-run effect	0.236***	0.251***	0.210***	0.059***
	(0.001)	(0.015)	(0.021)	(0.009)
Counterfactual mean	0.005	0.000	0.009	0.387
Observations	25,880	26,015	27,621	12,318
Panel C: Never-treated comparison group				
Long-run effect	0.217***	0.232***	0.195***	0.039***
	(0.004)	(0.016)	(0.017)	(0.007)
Counterfactual mean	0.025	0.028	0.034	0.390
Observations	4,839	4,844	5,164	2,186
Panel D: Liberal arts college comparison group				
Long-run effect	0.227***	0.240***	0.212***	0.059***
	(0.006)	(0.012)	(0.016)	(0.004)
Counterfactual mean	0.013	0.010	0.013	0.380
Observations	19,636	19,695	20,954	12,387

Notes: Table displays the estimated effect of the switch to coeducation on male share of freshmen (column 1), male share of undergraduate students (column 2), male share of degrees earned (column 3), and male share of faculty (column 4), estimated using equation 5. Freshman and undergraduate enrollment is available only beginning in 1968-69 school year. Faculty data available in selected years beginning in 1971. Each panel uses the specified pool of institutions to construct a comparison group and estimate a counterfactual trend in major choices, conditional on college selectivity and historical affiliation with the Catholic Church. In panel B, we additionally condition on school size, as measured by number of degrees granted, and the pre-reform trend in STEM choice among all students. Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. Standard errors are estimated using a block bootstrap with 1,000 replications that accounts for intracluster correlation at the institution level.

Table A4: Long-run effect of coeducation on the share of women choosing other majors

	(1) Art	(2) Business	(3) Education	(4) Health
Panel A: Not-yet-treated comparison group				
Long-run effect	-0.001	-0.017	-0.017	0.034
	(0.009)	(0.013)	(0.022)	(0.022)
Counterfactual mean	0.073	0.121	0.184	0.163
Observations	5,505	5,505	5,505	5,505
Panel B: All-college comparison group				
Long-run effect	0.006	-0.027**	0.008	0.053**
	(0.008)	(0.012)	(0.019)	(0.021)
Counterfactual mean	0.074	0.130	0.161	0.129
Observations	27,618	27,618	27,618	27,618
Panel C: Never-treated comparison group				
Long-run effect	0.003	-0.013	-0.011	0.016
	(0.012)	(0.012)	(0.025)	(0.026)
Counterfactual mean	0.069	0.116	0.178	0.180
Observations	5,164	5,164	5,164	5,164
Panel D: Liberal arts college comparison group				
Long-run effect	-0.005	-0.029***	-0.003	0.065***
	(0.008)	(0.011)	(0.016)	(0.019)
Counterfactual mean	0.079	0.136	0.171	0.124
Observations	20,954	20,954	20,954	20,954

Notes: Table displays the estimated effect of the switch to coeducation on male share of degrees earned (column 1) and graduating female students' choice of major (columns 2-5), estimated using $\hat{\beta}_{LR}$ from equation 5. Each panel uses the specified pool of institutions to construct a comparison group and estimate a counterfactual trend in major choices, conditional on college selectivity and historical affiliation with the Catholic Church. In panel B, we additionally condition on school size, as measured by number of degrees granted, and the trend in STEM choice among all students. Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. Standard errors are estimated using a block bootstrap with 1,000 replications that accounts for intracluster correlation at the institution level. Counterfactual mean is the share of women that would have chosen each major at treated schools if choices at those schools had followed trends at the comparison group of institutions.

Table A5: Long-run effect of coeducation on the share of women choosing other majors

	(1) Home ec	(2) Humanities	(3) Other	(4) Psychology	(5) Soc sci
Panel A: Not-yet-treated comparison group					
Long-run effect	0.001	0.001	0.001	0.011	0.022
	(0.006)	(0.014)	(0.017)	(0.007)	(0.015)
Counterfactual mean	0.030	0.095	0.039	0.058	0.126
Observations	5,505	5,505	5,505	5,505	5,505
Panel B: All-college comparison group					
Long-run effect	0.004	-0.007	0.005	0.001	-0.002
	(0.006)	(0.010)	(0.012)	(0.007)	(0.012)
Counterfactual mean	0.029	0.107	0.042	0.066	0.145
Observations	27,618	27,618	27,618	27,618	27,621
Panel C: Never-treated comparison group					
Long-run effect	0.001	-0.001	0.015	0.008	0.026**
	(0.006)	(0.015)	(0.012)	(0.007)	(0.013)
Counterfactual mean	0.030	0.098	0.026	0.061	0.122
Observations	5,164	5,164	5,164	5,164	5,164
Panel D: Liberal arts college comparison group					
Long-run effect	0.003	-0.006	0.004	0.004	0.009
	(0.005)	(0.010)	(0.012)	(0.008)	(0.011)
Counterfactual mean	0.028	0.102	0.036	0.066	0.138
Observations	20,954	20,954	20,954	20,954	20,954

Notes: Table displays the estimated effect of the switch to coeducation on graduating female students' choice of major, estimated using $\hat{\beta}_{LR}$ from equation 5. Each panel uses the specified pool of institutions to construct a comparison group and estimate a counterfactual trend in major choices, conditional on college selectivity and historical affiliation with the Catholic Church. In panel B, we additionally condition on school size, as measured by number of degrees granted, and the trend in STEM choice among all students. Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. Standard errors are estimated using a block bootstrap with 1,000 replications that accounts for intracluster correlation at the institution level. Counterfactual mean is the share of women that would have chosen each major at treated schools if choices at those schools had followed trends at the comparison group of institutions.

Table A6: Long-run effect of coeducation on women's high school GPA and ranking within class

	(1) GPA rank in class	(2) GPA rank in STEM
Panel A: Not-yet-treated comparison group		
Long-run effect	0.0388***	0.0552
	(0.0108)	(0.0785)
Counterfactual mean	0.391	0.336
Observations	1,426	1,363
Panel B: All-college comparison group		
Long-run effect	0.0463***	-0.0033
	(0.0116)	(0.0113)
Counterfactual mean	0.366	0.350
Observations	4,680	4,250
Panel C: Never-treated comparison group		
Long-run effect	0.0405***	0.0567
	(0.0114)	(0.0786)
Counterfactual mean	0.390	0.333
Observations	1,400	1,339
Panel D: Liberal arts college comparison group		
Long-run effect	0.0495***	-0.0152
	(0.0105)	(0.0123)
Counterfactual mean	0.371	0.362
Observations	9,084	8,639

Notes: Table displays the estimated effect of the switch to coeducation on female students' average high-school GPA ranking among their college freshman classmates (column 1), and female students' average high-school GPA ranking among college freshman classmates who intended to major in STEM (column 2), estimated using $\hat{\beta}_{LR}$ from equation 5. Each panel uses the specified pool of institutions to construct a comparison group and estimate a counterfactual trend in major choices, conditional on college selectivity and historical affiliation with the Catholic Church. In panel B, we additionally condition on school size, as measured by number of degrees granted. Data drawn from the HERI Freshman Survey, spanning 1966-2006, linked to hand-collected dates of transitions to coeducation by institution. Standard errors are estimated using a block bootstrap with 1,000 replications that accounts for intracluster correlation at the institution level.

Table A7: Bounding the composition effect of coeducation on STEM degree receipt, alternative comparison groups

	(1)	(2)	(3)	(4)
Panel A: Effect of freshman characteristics on women's likelihood of earning STEM degree				
Effect of intent to major in STEM	0.336***	0.333***	0.332***	0.317***
	(0.040)	(0.040)	(0.040)	(0.041)
Covariates:				
Career, family aspirations		X	X	X
Parental education, occupation			X	X
High school grades, coursework				X
R-squared	0.191	0.199	0.205	0.215
Observations	1,235	1,235	1,235	1,235
Panel B: Effect of coeducation on predicted share of female freshmen who will major in STEM, never-treated comparison group				
Estimated composition effect	0.006	0.010	0.010	0.014
	(0.009)	(0.011)	(0.011)	(0.012)
Composition effect / Total effect of coeducation on STEM major choice	-18%	-31%	-30%	-40%
Composition effect upper bound	37%	34%	33%	27%
Panel C: Effect of coeducation on predicted share of female freshmen who will major in STEM, liberal arts college comparison group				
Estimated composition effect	0.002	0.000	0.001	0.001
	(0.004)	(0.004)	(0.004)	(0.004)
Composition effect / Total effect of coeducation on STEM major choice	-5%	-1%	-2%	-3%
Composition effect upper bound	20%	24%	23%	21%

Notes: Panel A reports regression estimates of the effect of intention to major in STEM as of freshman year on share of students earning STEM degree, derived from sample of women in National Longitudinal Study of 1972. Panels B and C report implied long-run effect on the predicted share of freshman women at newly coeducational colleges who will major in STEM, calculated using equation 5 and sample of women from the HERI Freshman Survey. Predicted share in STEM is constructed by interacted coefficients from the regressions in panel A with characteristics of entering freshman women in the HERI data. Share of total effect explained by composition is constructed by dividing predicted STEM effect by estimated effect of coeducation on the share of women earning STEM degree from our linked IPEDS-HERI data (-0.034, see Figure A1a). Upper bound on composition effect is constructed by dividing lower bound of 95% confidence interval of predicted STEM effect by -0.034. See Appendix Table A7 for estimates drawing on alternative comparison groups.

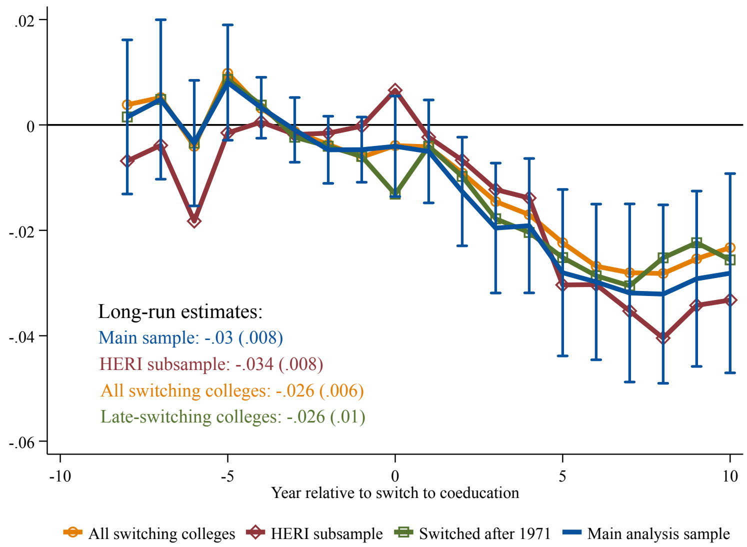
Table A8: Effect of coeducation on the share of women in the sophomore class majoring in STEM

	(1) Sophomore class	(2) Long-run effect
Panel A: Not-yet-treated comparison group		
Effect on STEM	-0.0126***	-0.0302***
	(0.0053)	(0.0077)
Effect / Δ male grads	-0.176	-0.157
Panel B: All-college comparison group		
Effect on STEM	-0.0123***	-0.0312***
	(0.0053)	(0.0067)
Effect / Δ male grads	-0.159	-0.148
Panel C: Never-treated comparison group		
Effect on STEM	-0.0153***	-0.0349***
	(0.0059)	(0.0092)
Effect / Δ male grads	-0.208	-0.179
Panel D: Liberal arts college comparison group		
Effect on STEM	-0.0162***	-0.0349***
	(0.0058)	(0.0059)
Effect / Δ male grads	-0.206	-0.164

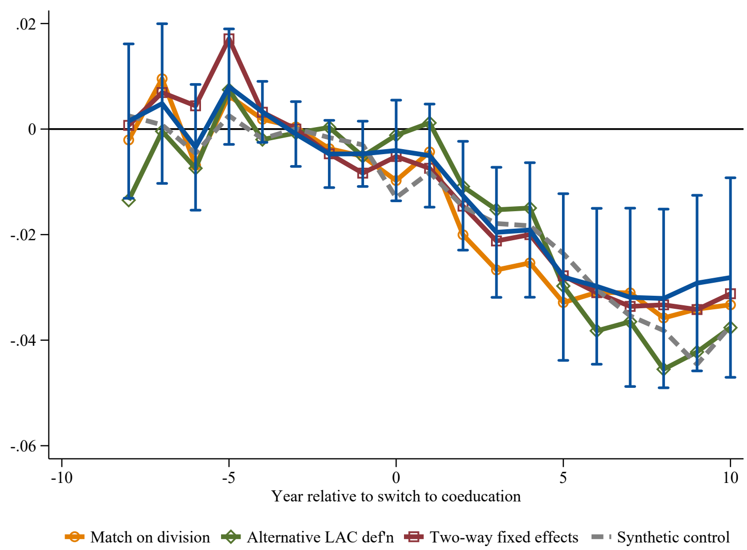
Notes: Table displays the estimated effect of the switch to coeducation on the share of women majoring in a STEM field. Estimates in column 1 correspond to $\hat{\beta}_{\tau=2}$ from equation 3, and estimates in column 2 correspond to equation 5. Each panel uses the specified pool of institutions to construct a comparison group and estimate a counterfactual trend in major choices, conditional on college selectivity and historical affiliation with the Catholic Church. In panel B, we additionally condition on school size, as measured by number of degrees granted, and the trend in STEM choice among all students. Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. Standard errors are estimated using a block bootstrap with 1,000 replications that accounts for intracluster correlation at the institution level. Third row of each panel rescales the estimated effect by the effect on male share of graduates in event-year $\tau = 2$ (column 1) or in event-years 5 through 9 (column 2).

Figure A1: Robustness of estimated effect of coeducation on share of women majoring in STEM to sample criteria

(a) Robustness to selection of treatment group

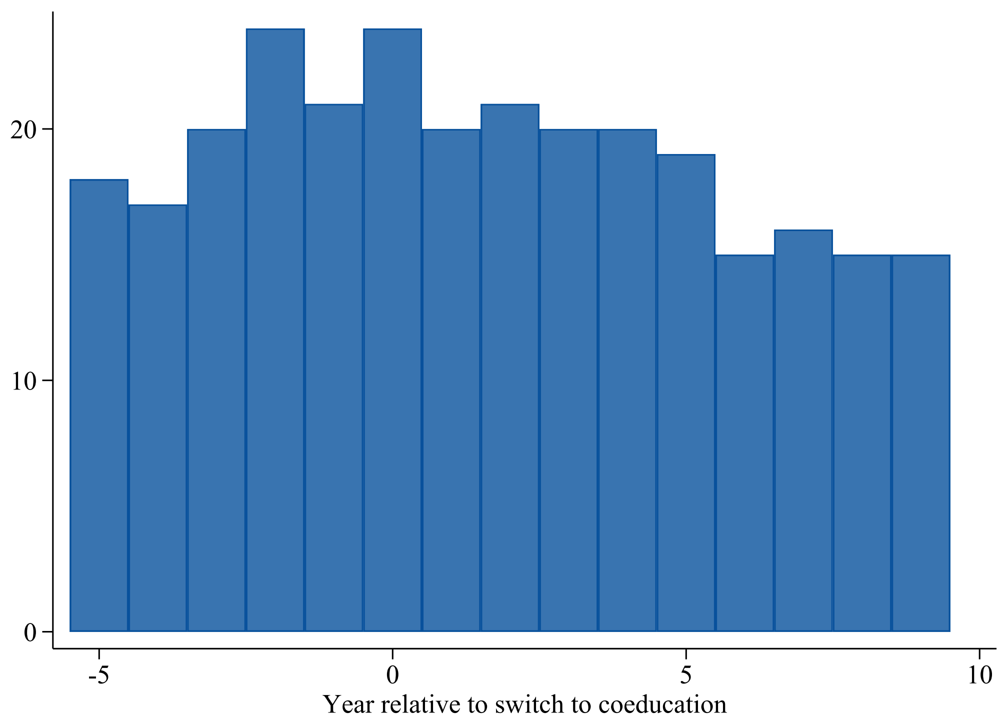


(b) Robustness to construction of comparison group



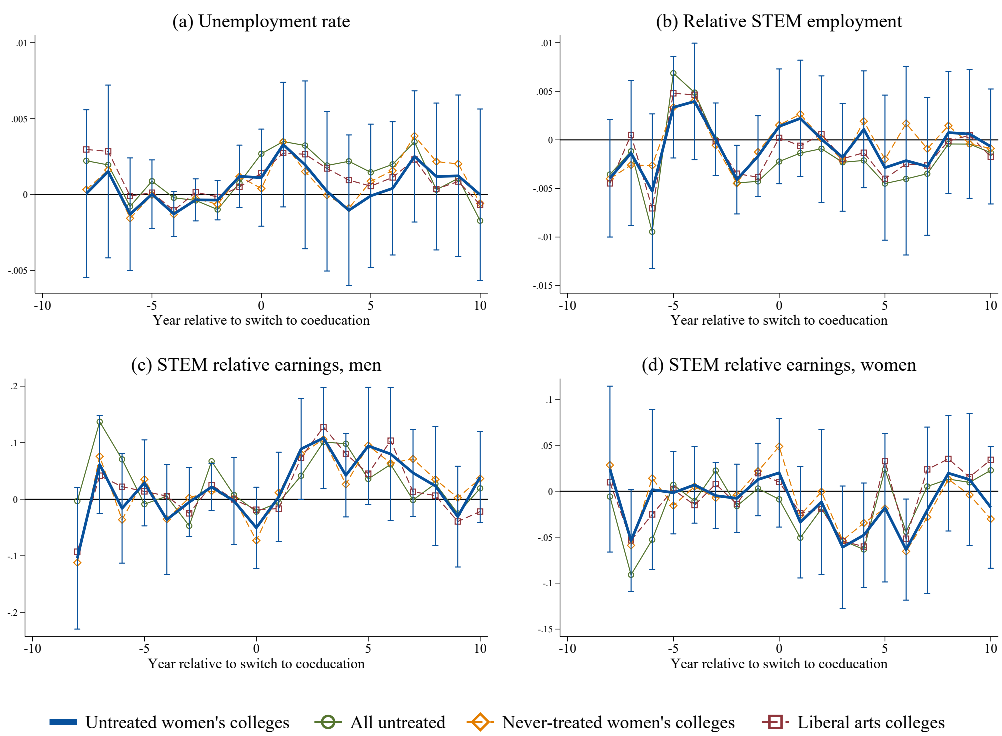
Notes: Figures show event-study estimates from equation 3. In each figure, blue line replicates estimates from Figure 3, which is based on a sample of 77 treated schools. Other lines show estimated effect on share of women in STEM using alternative treatment (Figure A1a) or comparison (Figure A1b) groups.

Figure A2: Representation of former women's colleges in The Freshman Survey



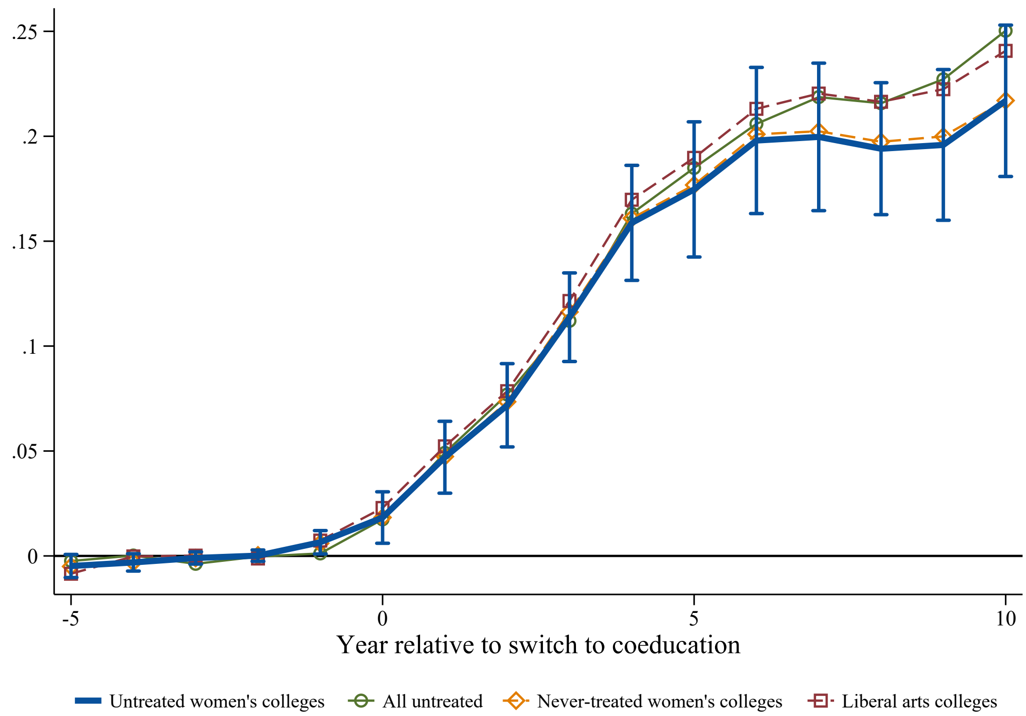
Notes: Data drawn from 1966-2006 versions of The Freshman Survey administered by HERI, linked to hand-collected dates of transitions to coeducation by institution's state. Each bar shows the number of treated schools that appear in the survey in each year relative to the switch to coeducation. Sample of treated schools is limited to 30 institutions that were surveyed at least once in the five years prior and once in the 10 years after the reform.

Figure A3: Tests for coinciding labor-market shocks



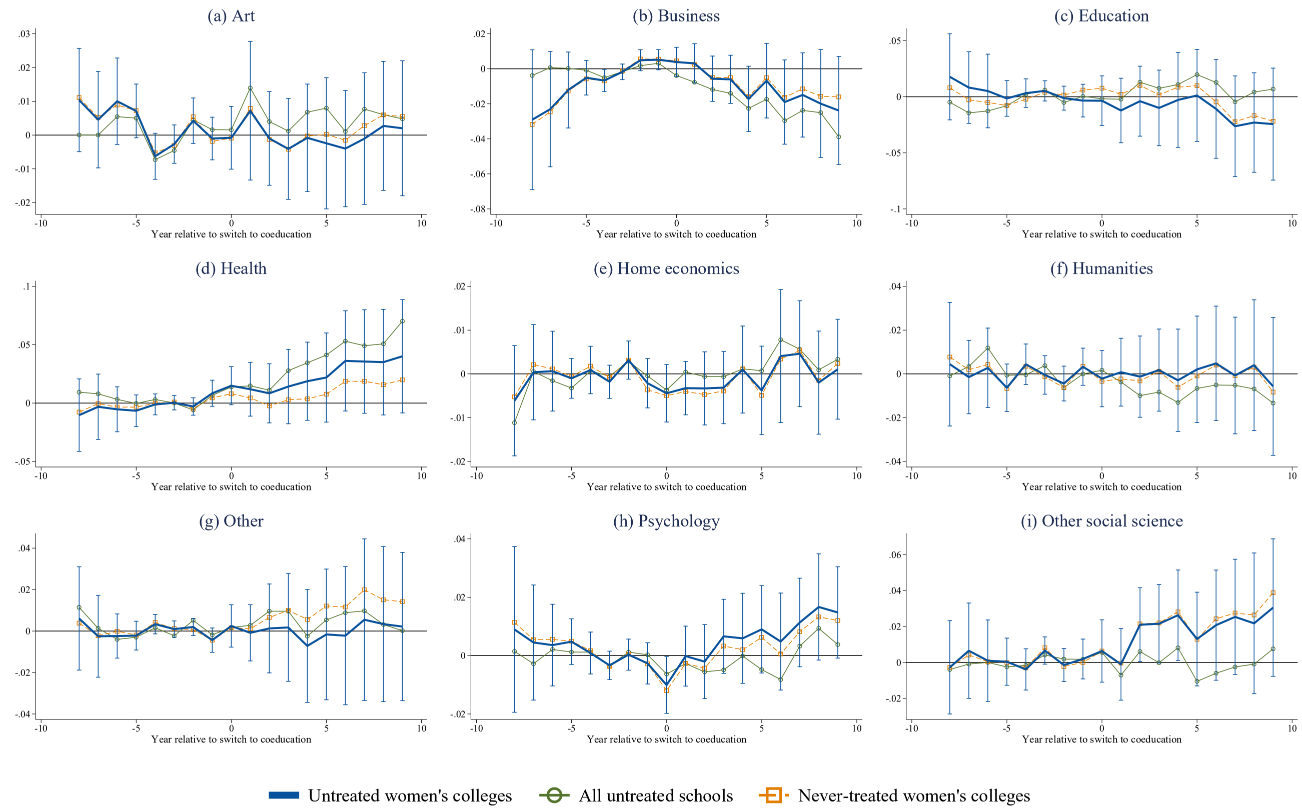
Notes: Data drawn from 1966-2016 CPS data accessed via IPUMS (Ruggles et al., 2020), linked to hand-collected dates of transitions to coeducation by institution's state. See Section 2 for further detail. Unemployment rate is measured among individuals age 18-64. Relative STEM employment is constructed as the ratio of college-educated workers in STEM occupations to workers in non-STEM occupations. Relative income among men is constructed as the ratio of average annual income among college-educated men currently working in a STEM occupation to average annual income among college-educated men currently working in a non-STEM occupation. Relative income for women in STEM is constructed in the same manner, except that we include individuals with 0 earnings in the previous year. Panels display estimates of β_k from equation 3. Standard errors are constructed from a block bootstrap clustered at the institution level.

Figure A4: Effect of coeducation on male share of graduates



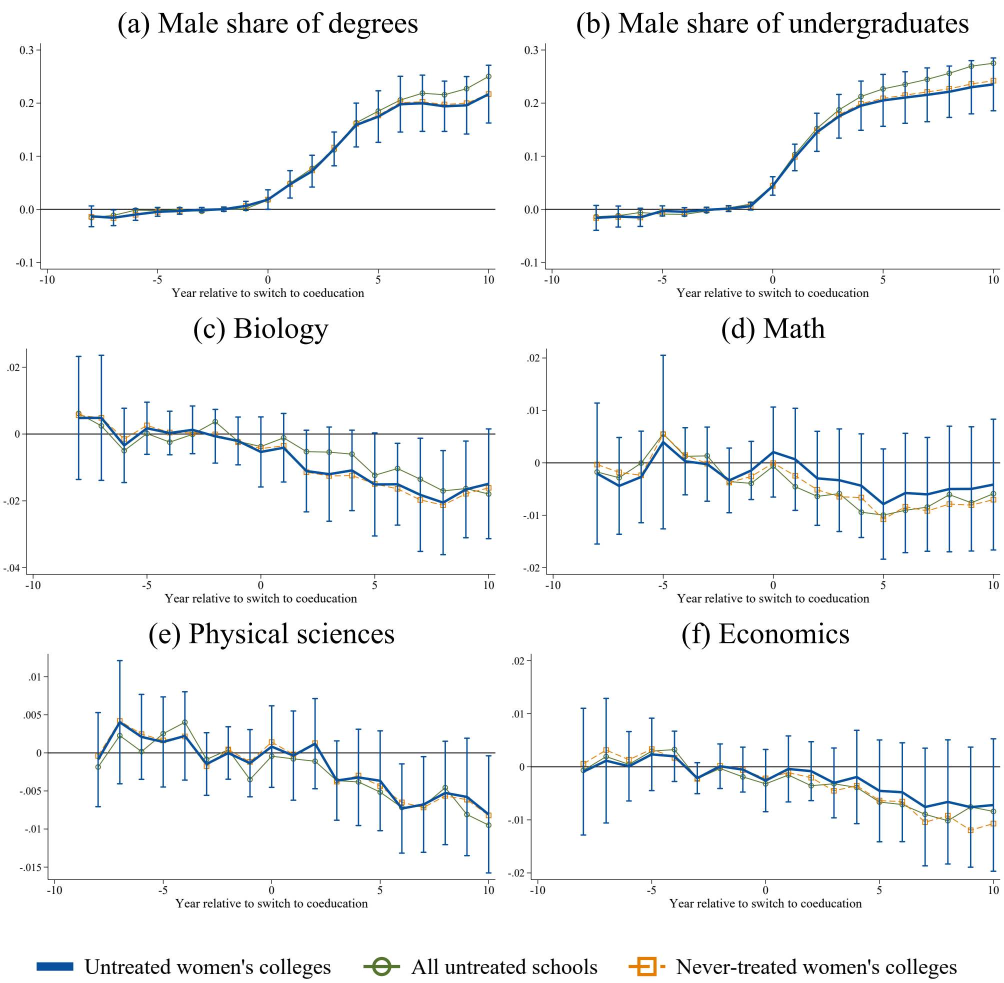
Notes: Figure shows event-study estimates from equation 3, using comparison group specified in legend. Data on degrees granted comes from HEGIS and IPEDS surveys, 1966-2016, linked to hand-collected information on dates of transition to coeducation. 95% confidence intervals are constructed using block bootstrap clustered at the institution level.

Figure A5: Effect of coeducation on share of women choosing other degrees



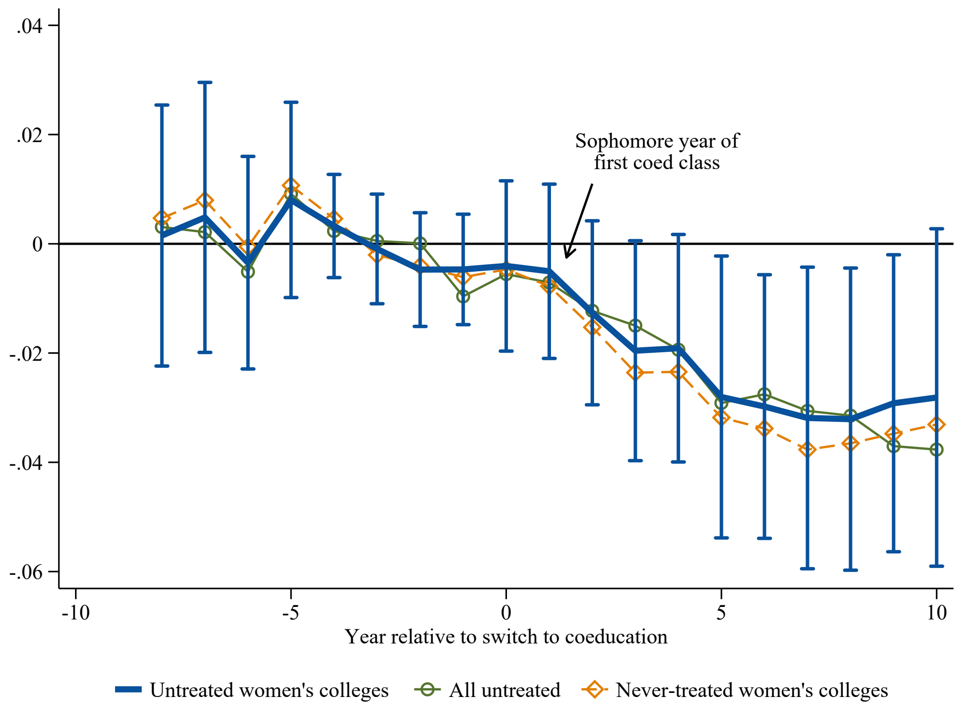
Notes: Figure shows event-study estimates from equation 12, aggregated using equation 3, using comparison group specified in legend. Data on degrees granted comes from HEGIS and IPEDS surveys, 1966-2016, linked to hand-collected information on dates of transition to coeducation. 95% confidence intervals are constructed using block bootstrap clustered at the institution level.

Figure A6: Effect of coeducation on gender composition and women's choice of quantitative majors, with uniform confidence intervals



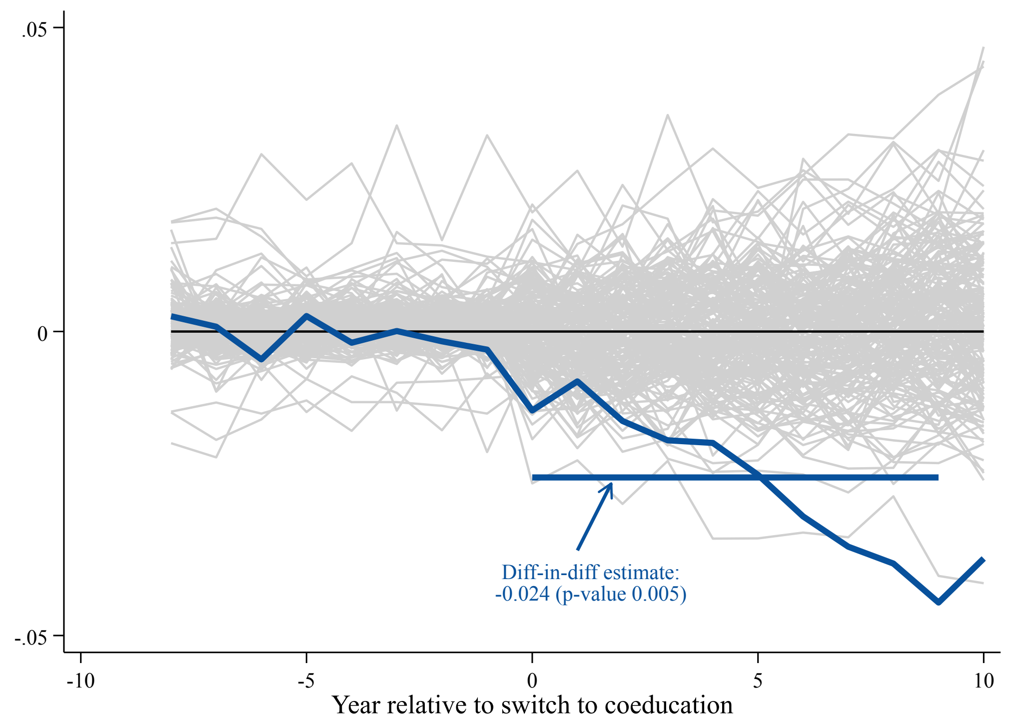
Notes: Figure repeats point estimates from Figure 4 with confidence intervals that account for multiple testing across event-time periods.Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. See Section 2 for further detail. Panels display estimate of β_k from equation 3. Dependent variable is the share of degrees earned in STEM among all degrees earned by women in the academic year.

Figure A7: Effect of coeducation on the share of women majoring in STEM, with uniform confidence intervals



Notes: Figure repeats point estimates from Figure 3 with confidence intervals that account for multiple testing across event-time periods.Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. See Section 2 for further detail. Panels display estimate of β_k from equation 3. Dependent variable is the share of degrees earned in STEM among all degrees earned by women in the academic year. STEM fields include math, biology, physical sciences, engineering, engineering technology, and computer science.

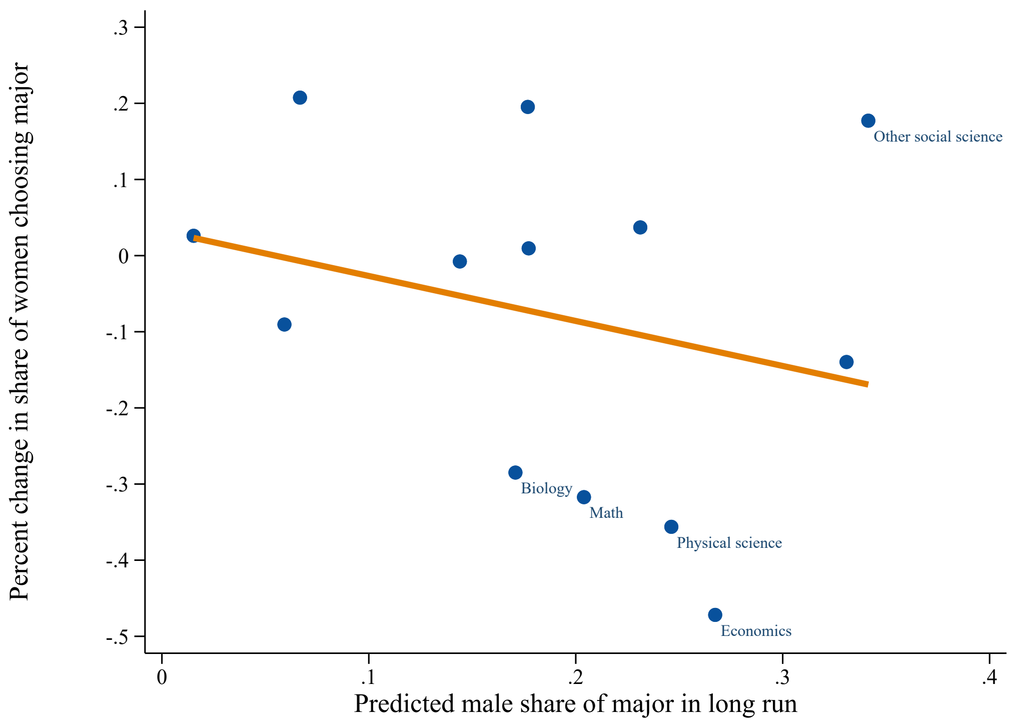
Figure A8: The effect of coeducation on the STEM share of degrees awarded to women: synthetic control specification



Notes: Data drawn from HEGIS/IPEDS surveys, spanning 1966-2016, linked to hand-collected dates of transitions to coeducation by institution. See Section 2 for further detail. The majors included in the STEM concentration are described in Section 2 and Appendix B. See Appendix D for description of the synthetic controls procedure. Dark line reports the main estimate, while grey lines report the results of a randomization inference procedure with 250 replications.

Figure A9: Effect of male inflows on distribution of women's choices of major

(a) Effect on women's choice of major vs. male inflow to major



(b) Elasticity of women's choice of major as share of total male inflow

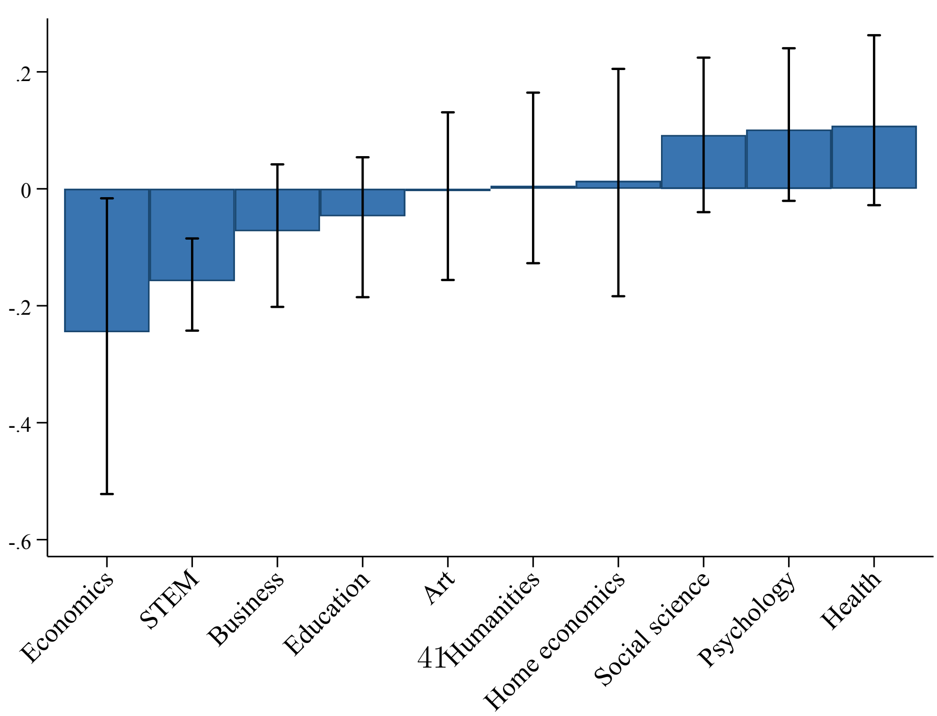
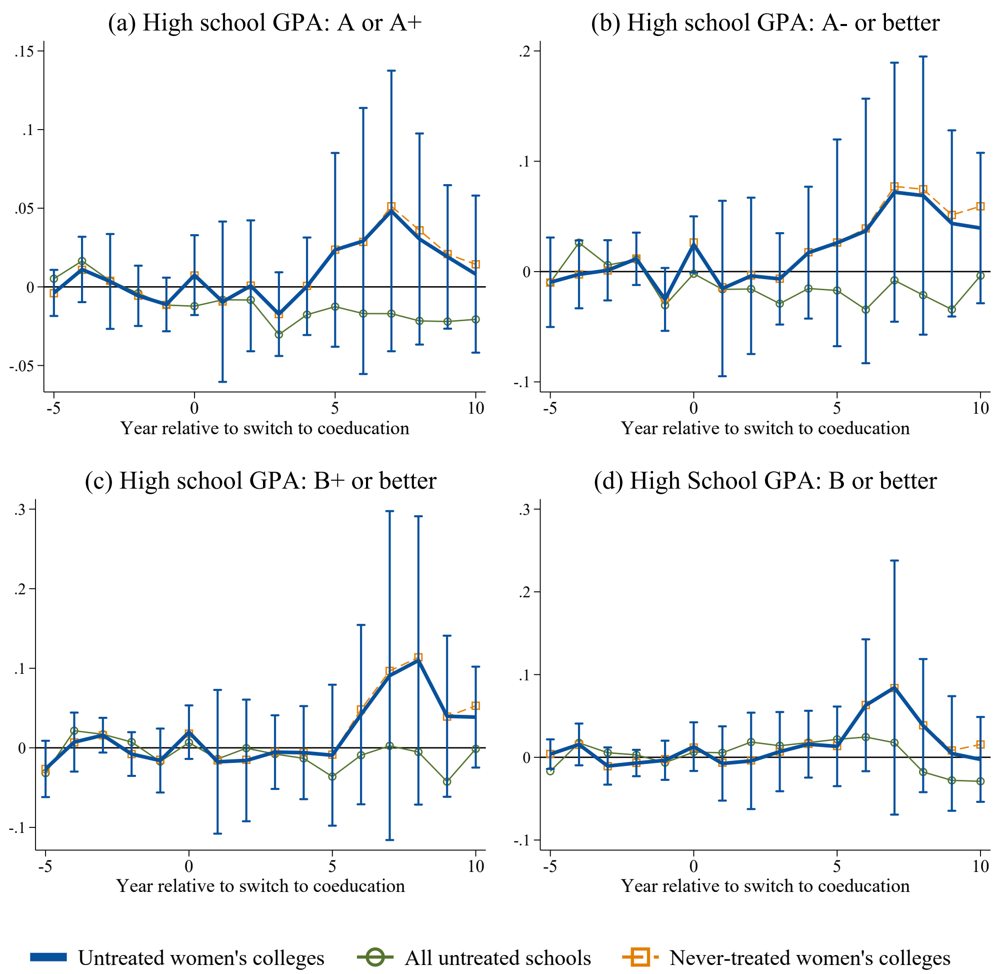


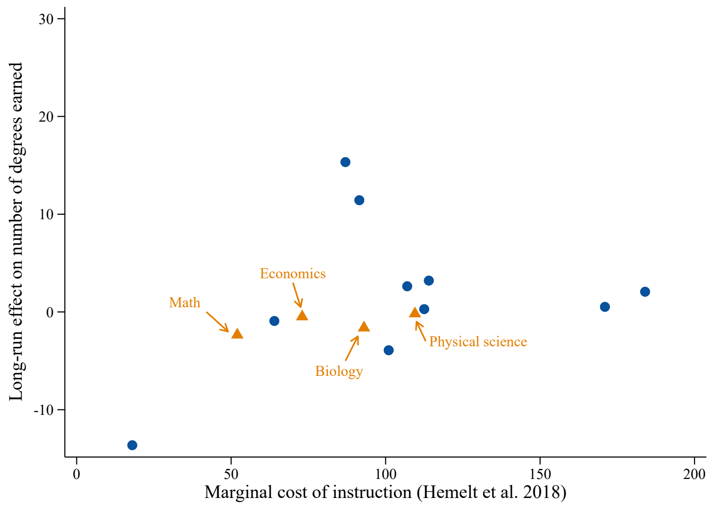
Figure A10: The effects of coeducation on the distribution of female matriculants' high school GPA



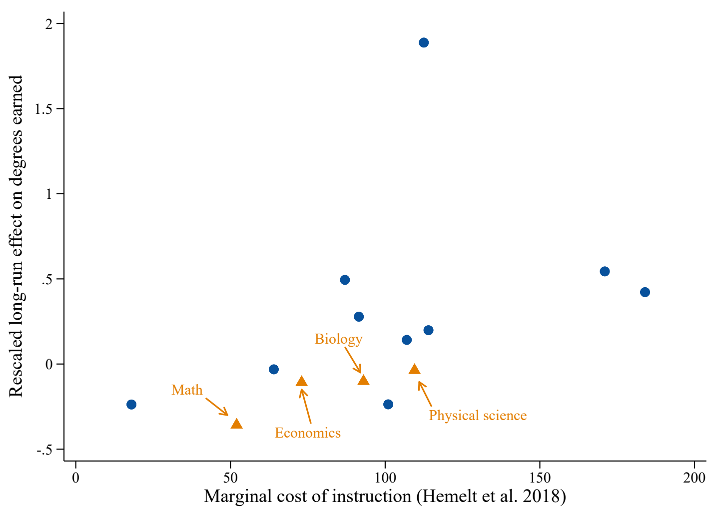
Notes: Each point shows and estimate of β_k from equation 3, using the comparison group specified in the legend. Error bars show 95% confidence interval constructed using block bootstrap clustered at the institution level. Data drawn from HERI, spanning 1966-2006, linked to hand-collected dates of transitions to coeducation by institution. See Section 2 for further detail.

Figure A11: Relationship between growth in total degrees earned and marginal cost of instruction by field

(a) Long-run effect vs marginal cost of instruction



(b) Rescaled long-run effect vs marginal cost of instruction



Notes: Long-run effects on degrees earned are estimated using equation 5 and data drawn from HEGIS/IPEDS surveys. In Figure A11b, effects are rescaled by the counterfactual mean of total degrees earned in the field. Marginal cost of instruction come comes from Hemelt et al. (2018) estimates among colleges with no graduation programs (see Table 5, column 5).