Supplementary Materials For “The Effects of Prohibiting Marriage Bars: The Case of U.S. Teachers”

Amy Kim
Princeton University*

Carolyn Tsao
Microsoft Research†

*Department of Economics, Princeton University, kimamy@princeton.edu

†Microsoft Research, Economics and Computation, New York carolyntsao@microsoft.com

A Additional Figures

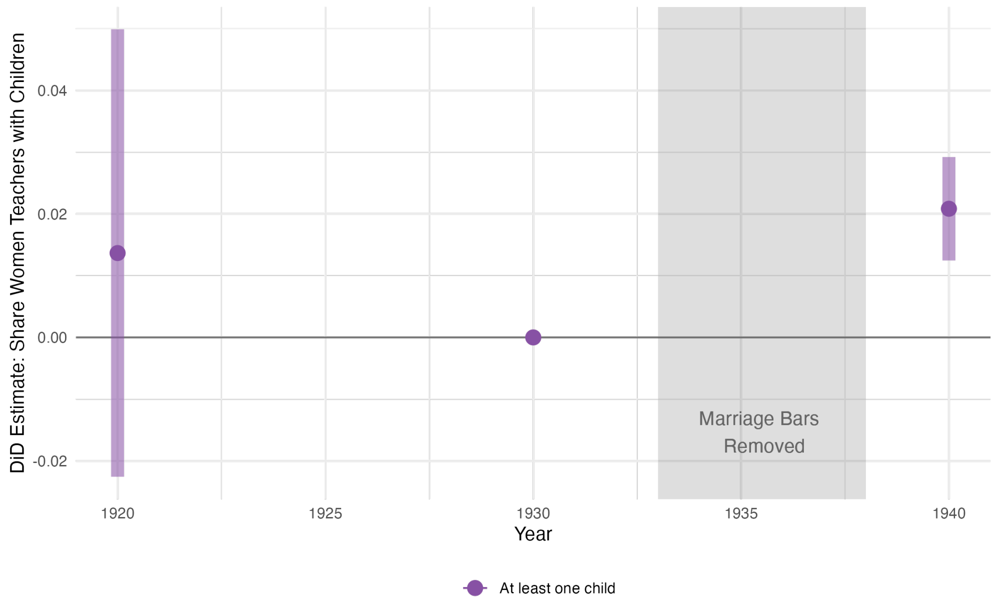


Figure A1: Results on fertility: Estimated effects of the introduction of employment protections for married women in teaching on the county shares of *white women teachers* with children.

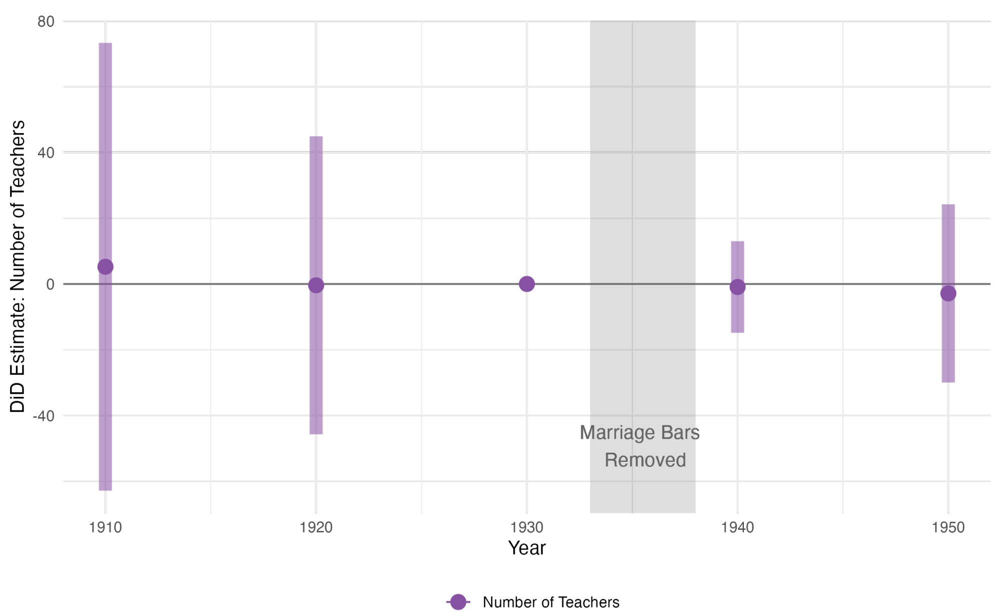


Figure A2: Estimated effects of the introduction of employment protections for married women on the total number of white teachers per county.

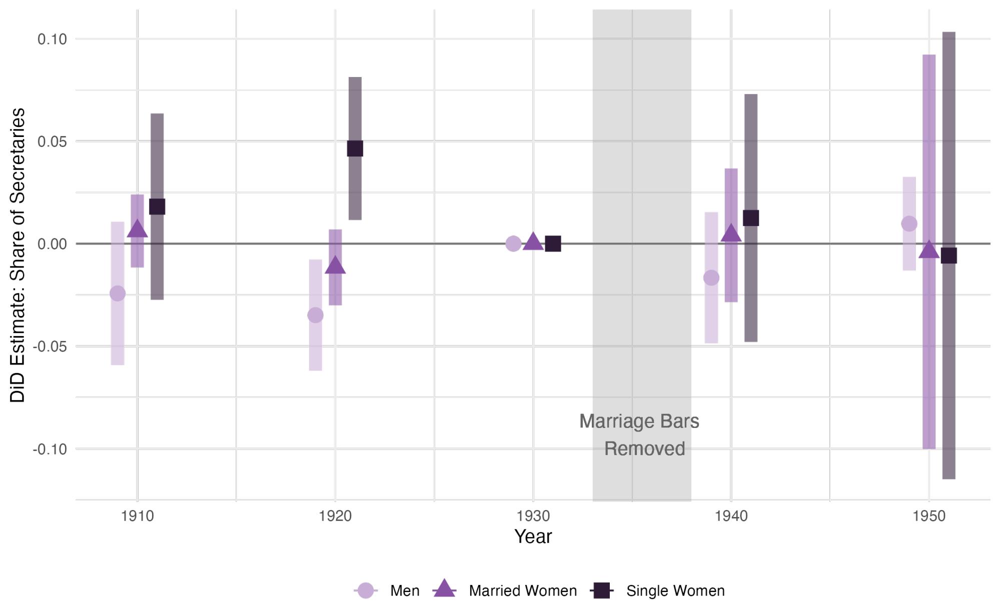


Figure A3: Placebo test: Estimated effects of the introduction of employment protections for married women in teaching on the county shares of *secretaries* who are men, unmarried women, and single women.

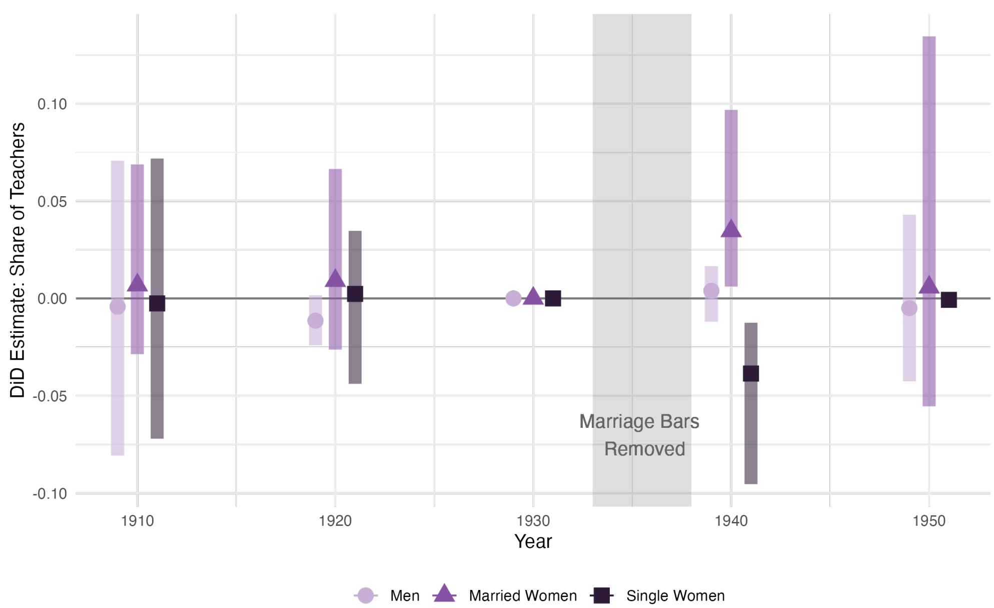


Figure A4: Main results, with state-level clustering of standard errors using wild cluster bootstrapping (Fischer and Roodman 2021): Estimated effects of the introduction of employment protections for married women in teaching on the county shares of *teachers* who are men, unmarried women, and single women.

B Additional Tables

Table B1: Census linkage rates by year and group.

	Treated States	Control States	White Women	Black Women	Unmarried Women	Married Women	Unmarried Women Teachers	Married Women Teachers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1910-1920 Linked Sample								
% 1910 Individuals Linked to 1920	60.9%	56.3%	61.0%	30.6%	53.9%	62.8%	47.9%	59.8%
Number of 1910 Individuals (Thous.)	4,500.8	6,988.0	39,530.1	5,018.5	27,002.2	17,689.2	415.4	24.6
1920-1930 Linked Sample								
% 1920 Individuals Linked to 1930	62.2%	57.4%	62.0%	32.1%	53.9%	65.8%	47.8%	66.4%
Number of 1920 Individuals (Thous.)	4,978.4	7,798.1	46,387.0	5,276.4	30,334.6	21,490.5	538.3	58.3
1930-1940 Linked Sample								
% 1930 Individuals Linked to 1940	62.8%	59.4%	62.5%	35.3%	53.0%	68.5%	45.6%	71.4%
Number of 1930 Individuals (Thous.)	5,784.9	8,510.2	54,423.5	6,035.6	34,453.1	26,242.3	679.0	150.5

Notes: Linkage rates are computed as the share of a given population in the base year (for example 1910) that are successfully linked to the following census (for example 1920). Groups (for example marital status) are based on base year characteristics. Treated states are NC and KY. We use Census Tree linkages from 1910-1920, 1920-1930, and 1930-1940 (Price et al. 2023a,b,c).

Table B2: Summary statistics on retail sales per capita during the Great Depression, by treated status

	1929 Retail Sales per cap (1967 USD) (1)	1929-33 RS per cap growth (%) (2)	1933-39 RS per cap growth (%) (3)	1939 RS per cap (1967 USD) (4)
Treated Counties				
Kentucky	441.8	-41.2	41.5	443.0
North Carolina	406.9	-35.8	41.6	431.1
Treated Average	422.7	-38.4	41.5	436.4
Neighboring Southern Control Counties				
South Carolina	339.7	-23.5	45.6	423.9
Tennessee	485.2	-43.4	47.5	505.3
Virginia	488.8	-28.0	43.3	569.3
West Virginia	510.0	-36.7	37.8	515.3
Control Average	461.5	-34.0	43.8	509.1

Source: Fishback, Horrace, and Kantor (2005).

Table B3: Heterogeneity in estimated effects of prohibitions on the share of teachers that are married women.

Dependent Variable:	Pr(Married Woman — Teacher)								
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variables									
Treated $\times 1940$ (γ^{DD}_{1940})	0.0350***	0.0347***	0.0240***	0.0313***	0.0375***	0.0388***	0.0346***	0.0405***	0.0397***
	(0.0073)	(0.0066)	(0.0079)	(0.0064)	(0.0075)	(0.0072)	(0.0072)	(0.0077)	(0.0068)
Treated $\times 1940 \times Nonwhite$						-0.0332**		-0.0426**	
						(0.0129)		(0.0171)	
Treated $\times 1940 \times Urban$							-0.0023	-0.0107	
							(0.0112)	(0.0121)	
Treated $\times 1940 \times Nonwhite \times Urban$								0.0306	
								(0.0232)	
Treated $\times 1940 \times (Age - 30)$									-0.0025***
									(0.0005)
Inverse Weighted by County	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	437,111	437,111	322,511	221,846	343,350	437,111	437,111	437,111	437,111
Adjusted R^2	0.09100	0.10761	0.09672	0.03215	0.10726	0.12169	0.10846	0.12249	0.12136

Notes: Estimation of Equation (2) for the cross-sectional sample of all teachers in our balanced sample of counties in treated (KY, NC) and neighboring Southern control states (VA, SC, TN, WV). The outcome variable for all regressions is an indicator for whether an individual is a married woman in t. Column (1) includes all teachers with no weights or controls. Column (2) includes inverse weighting of each observation by the number of teachers in a county and year, thus replicating the county-level regression in Table 2, Column (1). All remaining columns use county-year inverse population weights for comparability with our main county-level results. Column (3) restricts the sample to the set of counties with at least ten Black teachers in 1930 and 1940 (see the County Sample Selection Section of the main text for more details). Column (4) includes county-level controls for the share of workers in manufacturing, the share of workers in agriculture, unemployment rate, and log population. Note that the sample size is smaller because full-count Census data on employment status was only available in 1930 and 1940. Column (5) excludes counties in the Tennessee Valley Authority (see the Identifying Assumption Section of the main text for further discussion). Columns (6) and (7) include indicators for whether a teacher is nonwhite ($Nonwhite$) and whether a teacher lives in an urban area ($Urban$) respectively as additional interaction terms. Column (8) includes $Nonwhite$, $Urban$, and the interaction between the two as additional interaction terms. Column (9) includes an individual's age relative to 30 (the mean age for teachers in treated counties in 1930) as an additional interaction term. All regressions include county and year fixed effects and use 1910-1950 full-count cross-sectional decennial Census data unless otherwise stated (Ruggles et al. 2024). Standard errors are clustered at the county level.

Table B4: Estimated effects of the prohibitions on women's propensity to get married and to teach, work outside of teaching, and exit the labor force.

Dependent Variable:	Pr(Married in t)	Pr(Married Teacher in t)	Pr(Married Non-Teacher in LF in t)	Pr(Married Not in LF in t)
	(1)	(2)	(3)	(4)
Sample 4: Women who were married and working as non-teachers in $t-10$				
Treated × 1940 (γ^{DD}_{1940})	-0.0016	0.0002	0.0189	-0.0207
	(0.0049)	(0.0006)	(0.0148)	(0.0165)
Dep. Var. 1930 Treated Mean	0.8779	0.0032	0.2078	0.6669
Observations	248,533	248,533	248,533	248,533
Adjusted R^2	0.03724	0.00187	0.10338	0.09394
Sample 5: Women who were unmarried and working as non-teachers in $t-10$				
Treated × 1940 (γ^{DD}_{1940})	0.0044	0.0003	0.0145**	-0.0104
	(0.0043)	(0.0005)	(0.0072)	(0.0078)
Dep. Var. 1930 Treated Mean	0.5132	0.0021	0.0824	0.4287
Observations	320,304	320,304	320,304	320,304
Adjusted R^2	0.17755	0.00154	0.04312	0.15284

Notes: Estimation uses a difference-in-differences approach as outlined in the Empirical Strategy Section. To construct our estimation samples, we identify women in treated states (KY, NC) and neighboring Southern states (VA, SC, TN, WV) in 1920, 1930, and 1940 whom we are able to link over consecutive Census years (between 1910 and 1920, 1920 and 1930, or between 1930 and 1940) using Census Tree linkages. From these women, we construct the following samples: Sample 4 contains linked women who were aged 18-50, white, married and in the labor force but not working as teachers in 1920 and 1930. Sample 5 contains linked women who were aged 8-40, white, unmarried and in the labor force but not working as teachers in 1920 and 1930. All regressions are at the individual-year level, include year, county, and age fixed effects and are clustered at the county level.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Source: All regressions use the 1910-1920, 1920-1930, and 1930-1940 linked full-count Census samples from IPUMS and Census Tree.

Table B5: Estimated effects of marriage bar prohibitions on fertility, occupational scores, LFP, and mobility for women who were unmarried and teaching in $t-10$

Dependent Variable:	Has child in t	Occupational Score in t		In Labor Force in t		Moves state in t	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated × 1940	-0.0251**	0.8418**	-0.0814	0.0266**	-0.0033	0.0391***	0.0589***
	(0.0118)	(0.3655)	(0.2838)	(0.0128)	(0.0100)	(0.0095)	(0.0119)
Treated × 1940 × Married in t			1.526***		0.0442***		-0.0429***
			(0.4302)		(0.0151)		(0.0131)
Dep. Var. 1930 Treated Mean	0.4819	10.39	10.39	0.4091	0.4091	0.1421	0.1421
County fixed effects	Yes	Yes	Yes	Yes	Yes		
Age fixed effects	Yes	Yes	Yes	Yes	Yes		
Observations	59,542	59,542	59,542	59,542	59,542		
Adjusted R^2	0.10860	0.13437	0.46280	0.11917	0.47197	0.02193	0.04099

Notes: Estimation uses a difference-in-differences approach as outlined in the Empirical Strategy Section. All columns show results for women in linked Sample 1, defined in the notes of Table 3. The outcome in Column (1) is an indicator for whether the woman has child in year t. The outcome in Column (2) is the woman's occupational score in year t. We include women who are unemployed, for whom their occupational score is 0. The outcome in Column (4) is an indicator for whether the woman is in the labor force in t. The outcome in Column (6) is an indicator for whether the woman lives in a different state in year t compared to year $t-10$. Columns (3), (5), and (7) add to the previous column an interaction term for whether the woman is married in year t. All regressions are at the individual-year level, include year, county, and age fixed effects and are clustered at the county level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Source: All regressions use the 1910-1920, 1920-1930, and 1930-1940 linked full-count Census samples from IPUMS and Census Tree.

Table B6: Heterogeneity in estimated effects of the prohibitions on the propensity of women not in the labor force in $t-10$ to work as a non-teacher in t.

Dependent Variable:	Pr(Married Non-Teacher in LF in t)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Sample 2: Women who were married and not in the labor force in $t-10$						
Treated × 1940	0.0064*	0.0028	0.0071**	0.0077**	0.0048	0.0059
	(0.0035)	(0.0028)	(0.0033)	(0.0032)	(0.0039)	(0.0062)
Observations	3,125,563	3,125,563	3,110,334	2,279,708	2,332,161	914,032
Adjusted R^2	0.01862	0.01560	0.01875	0.02253	0.02014	0.02101
Sample 3: Women who were unmarried and not in the labor force in $t-10$						
Treated × 1940	0.0044**	0.0020	0.0045**	0.0051***	0.0041*	0.0005
	(0.0021)	(0.0017)	(0.0021)	(0.0019)	(0.0024)	(0.0034)
Observations	2,217,852	2,217,852	2,209,789	1,511,858	1,688,568	646,957
Adjusted R^2	0.02051	0.01699	0.02064	0.02383	0.02206	0.02018
0.02022	0.01997					

Notes: Estimation uses a difference-in-differences approach as outlined in the Empirical Strategy Section. All columns show results for women in Samples 2 and 3 as defined in the notes of Table 3. Column (1) shows the unweighted estimates, replicating the main result in Table 3, Column (3). Column (2) includes inverse weighting of each observation by the number of observations in a county and year, imitating a county-level regression. Column (3) includes controls for industry share and population. Column (4) adds controls for county-level unemployment; note that the sample is smaller because the variable on employment status is not available in the 1920 Census. Column (5) excludes counties that were affected by the Tennessee Valley Authority (TVA). Column (6) includes only border counties, defined in Online Appendix C. All regressions are at the individual-year level, include year, county, and age fixed effects and are clustered at the county level. $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Source: All regressions use the 1910-1920, 1920-1930, and 1930-1940 linked full-count Census samples from IPUMS and Census Tree.

Table B7: Estimated effects of the prohibitions on women's propensity to remain unmarried and to teach, work outside of teaching, and exit the labor force.

Dependent Variable:	Pr(Unmarried in t)	Pr(Unmarried Teacher in t)	Pr(Unmarried Non-Teacher in LF in t)	Pr(Unmarried Not in LF in t)
	(1)	(2)	(3)	(4)
Sample 3: Women who were unmarried and not in the labor force in $t-10$				
Treated × 1940 (γ_{1940}^{DD})	0.0003	0.0014	-0.0058	0.0047
	(0.0036)	(0.0009)	(0.0036)	(0.0039)
Dep. Var. 1930 Mean	0.5224	0.0323	0.1918	0.2982
Observations	2,217,852	2,217,852	2,217,852	2,217,852
Adjusted R^2	0.15575	0.01056	0.06495	0.14088
Sample 5: Women who were unmarried and working as non-teachers in $t-10$				
Treated × 1940 (γ_{1940}^{DD})	-0.0044	0.0014	-0.0047	-0.0011
	(0.0043)	(0.0009)	(0.0064)	(0.0061)
Dep. Var. 1930 Mean	0.4868	0.0096	0.3498	0.1274
Observations	320,304	320,304	320,304	320,304
Adjusted R^2	0.17755	0.00284	0.12281	0.07548

Notes: See notes for Tables 3 and Online Appendix Table B6.

Table B8: Estimated effects of prohibitions on the labor force participation of women.

Dependent Variable:	Pr(In Labor Force)							
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
Treated $\times 1940$ (γ^{DD}_{1940})	0.0062	0.0091**	0.0001	0.0042	0.0019	0.0083*	0.0057	0.0095
	(0.0046)	(0.0039)	(0.0087)	(0.0045)	(0.0057)	(0.0049)	(0.0054)	(0.0069)
Treated $\times 1940 \times Married$			0.0072					
			(0.0108)					
Treated $\times 1940 \times Nonwhite$				-0.0023				
				(0.0132)				
Treated $\times 1940 \times Urban$					0.0141			
					(0.0118)			
Inverse Weighted by County	No	Yes	No	No	No	No	No	No
Observations	21,423,773	21,423,773	21,423,773	21,423,773	21,423,773	16,739,277	16,533,201	5,859,581
Adjusted R^2	0.07115	0.06639	0.16767	0.10433	0.08549	0.07904	0.07183	0.07169

Notes: Estimation of Equation (2) for the cross-sectional sample of all women in our balanced sample of counties in treated (KY, NC) and neighboring Southern control states (VA, SC, TN, WV). The outcome variable for all regressions is an indicator for whether an individual is in the labor force. Column (1) includes all women with no weights or controls. Column (2) includes inverse weighting of each observation by the number of women in a county and year, effectively imitating a county-level regression. All remaining columns do not use weights. Columns (3), (4) and (5) include indicators for whether a woman is married ($Married$), whether a woman is non-white ($Nonwhite$), and whether a woman lives in an urban area ($Urban$) respectively as additional interaction terms. Column (6) includes county-level controls for the share of workers in manufacturing, the share of workers in agriculture, unemployment rate, and log population. Note that the sample size is smaller because full-count Census data on employment status was only available in 1930 and 1940. Column (7) excludes counties in the Tennessee Valley Authority. Column (8) includes only border counties, defined in Online Appendix C. All regressions include county, year, and age fixed effects and use 1910-1950 full-count cross-sectional decennial Census data unless otherwise stated (Ruggles et al. 2024). Standard errors are clustered at the county level.

C Matched and Border Counties Designs

Our preferred specification relies on the assumption that in the absence of the laws passed in North Carolina and Kentucky, the composition of the teaching workforce would have evolved similarly in the treated states and neighboring southern states of South Carolina, Tennessee, Virginia, and West Virginia. To test whether our results are robust to alternative specifications, we employ two alternate empirical designs. The first is a border county design which narrows the sample of counties to treatment and control counties that are geographically adjacent and therefore more likely to be similar and satisfy the parallel trends assumption. The second is a matched counties design which uses data on counties across the country and does not rely on the assumption that geographically close counties are similar.

Border Counties Design

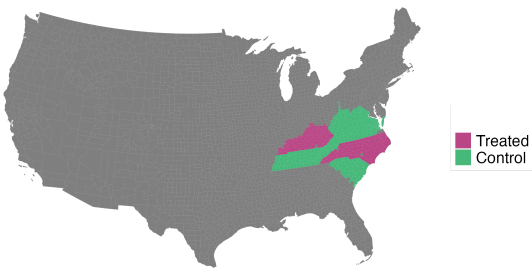
The reasoning underlying the border counties design is that counties in different states that border each other are highly likely to have followed similar historical trends, such that the only difference between counties on opposite sides of a border is their exposure to their states' state-wide prohibitions. Following this logic, we construct our alternate border county samples as follows. For our sample of treated border counties, we select all counties in treated states that border some other (non-treated) state. For our sample of control border counties, we select all counties in neighboring (non-treated) states that border a treated state.[1] Panel (b) of Figure C1 highlights the resulting samples of bordering counties.

Matched Counties Design

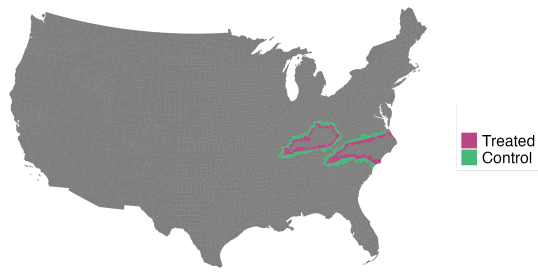
We match treatment and control counties using both the 1930 level and change between 1920 and 1930 of an extensive set of county-level variables, including demographics, urbanization, literacy rate, and workforce composition both for teachers and overall, all obtained from the full-count census (Ruggles et al. 2024). We also include 1929 retail sales per capita and the growth in retail sales per capita from 1929 to 1934 and from 1934 to 1939, as obtained from

[1]Note that for our border counties design, unlike our preferred specification, we do not require control counties to be in a Southern state.

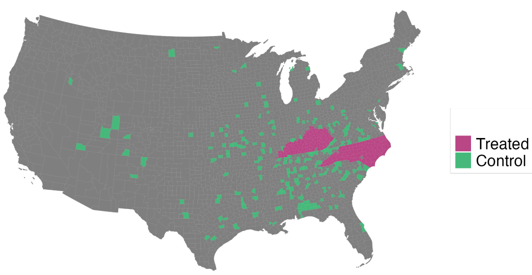
Figure C1: Maps of treated (pink) and control (green) counties for (a) our preferred control specification of neighboring Southern counties, (b) our border county specification and (c)-(e) specifications using a range of matching techniques.



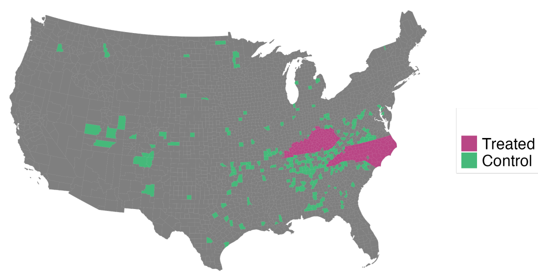
(a) Southern Neighbor States



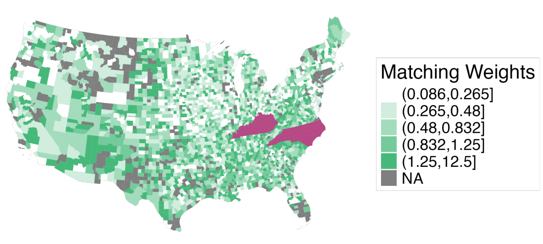
(b) Border Counties



(c) Match 1: Nearest Neighbor



(d) Match 2: Genetic



(e) Match 3: Optimal Full Matching

Fishback, Horrace, and Kantor (2005).[2]

To match counties we use three different methods, all utilizing the MatchIt package in R (Ho et al. 2011). The first matched sample is constructed by nearest neighbor matching using Mahalanobis distance; the second using genetic matching as developed by Diamond and Sekhon (2013) and Sekhon (2011); the third using optimal full matching as developed by Hansen (2004). The first two methods are one-to-one matching methods, which produce the same number of control counties as treatment counties. The third method, optimal full matching, uses all counties and assigns weights to control counties based on their similarity to treatment counties.

Figure C1 compares the control counties selected by the various matching methods to the neighboring southern states in our preferred specification. Matched samples 1 and 2 are geographically concentrated in the neighboring Southern states, reinforcing the fact that the neighboring Southern counties are indeed similar to our treated counties. Panel (d) of Figure C1 maps the weights of the control counties as determined by the optimal full matching method, which are not as closely concentrated in the neighboring states as with the other matching methods.

[2] Complete variable list (all at the county level): 1930 level and 1920-1930 change in population, share living in urban areas, share under age 20, share aged 20-39, share aged 40-59, share aged 60 or older, share white, share literate, share of 16-64-year-olds in the labor force, share of 16-64-year-old married women in the labor force, share of the labor force employed in manufacturing, share of the labor force employed in agriculture, share of teachers that are unmarried women, and share of teachers that are married women. 1929 level of retail sales per capita, 1929-1934 and 1934-1939 growth in retail sales per capita, and 1930 share of the labor force unemployed (all dollar amounts in 1967 USD). 1920-1930 change is calculated as $g_x = \frac{x_{1930} - x_{1920}}{x_{1920}}$, where x_t represents the value of the relevant variable x in year t, except for 1920-1930 change in share living in urban areas and share of teachers that are unmarried/married women, which are calculated as $g_x = \frac{x_{1930} - x_{1920}}{x_{1920} + 0.01}$ to avoid division by zero.

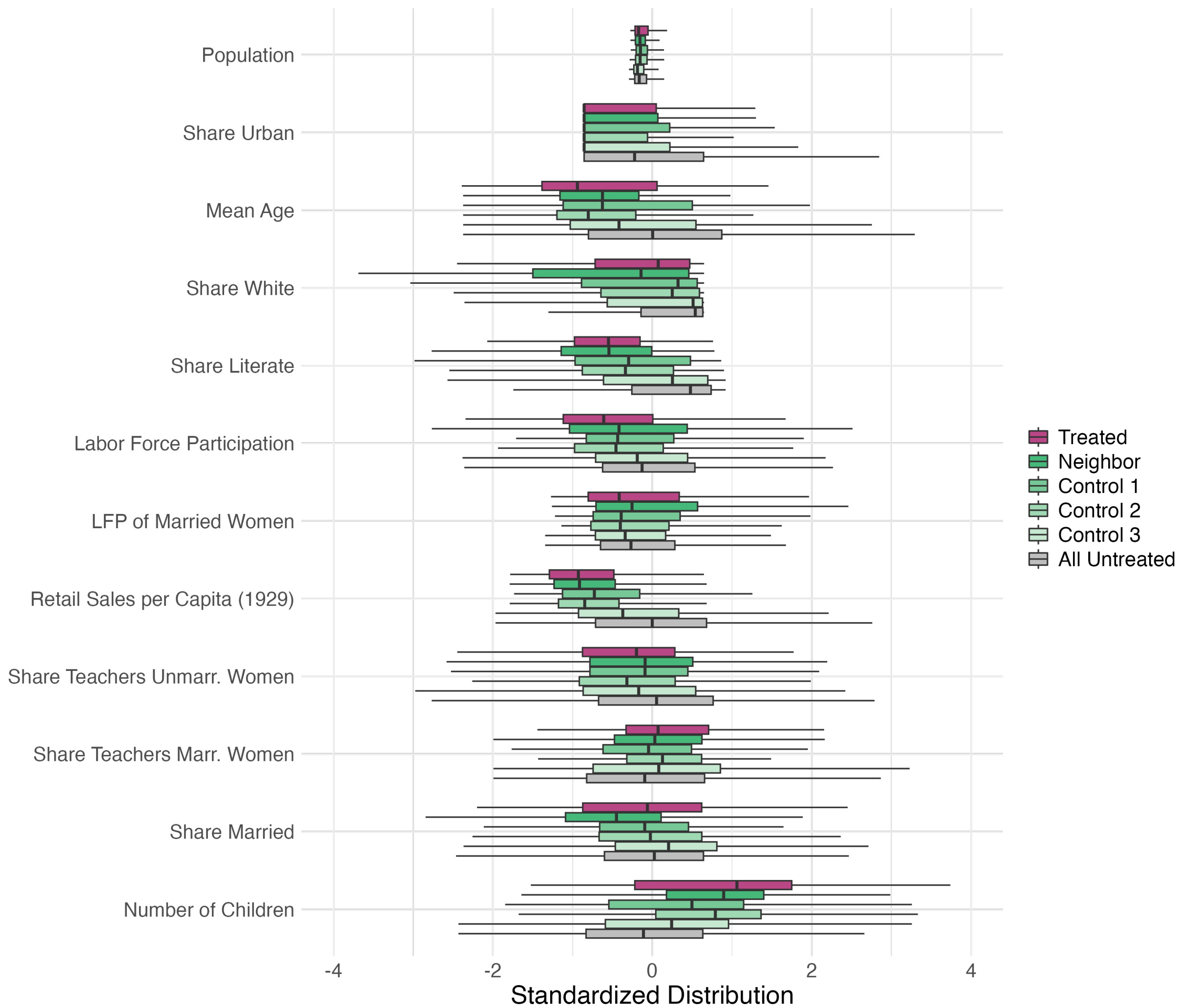


Figure C2: Boxplot of standardized 1930 values of various covariates by treatment or control group. The center bar represents the median, the edges of the box represent the 25th and 75th percentiles, and the edges of the whiskers represent extrema, with outliers removed (see R function geom_boxplot for further details). Distributions are weighted for control group 3. Covariates are outlined in detail in Footnote 2, and also include the share of women over the age of 18 that are married, and the average number of children for married women. All data is obtained from Ruggles et al. (2024) with the exception of 1929 Retail Sales per Capita, which is obtained from Fishback, Horrace, and Kantor (2005).

In Figure C2, we graph boxplots for the treated and various control groups of the standardized 1930 values of some of the matching covariates listed in Footnote 2, as well as two additional variables not used for matching (share of women married and average number of children for married women). Motivating the need to identify an appropriate control group for KY and NC, the boxplots show that the average untreated county is quite distinct from the treated counties. Importantly, the neighoring Southern counties are very similar to the treated counties, and on some dimensions (such as on the share of teachers married women or the number of children) even outperform the matched county groups in terms of similarity. While the first and second control groups are very similar in distribution to the treatment counties across nearly all covariates, the third control group is much less similar.

Results

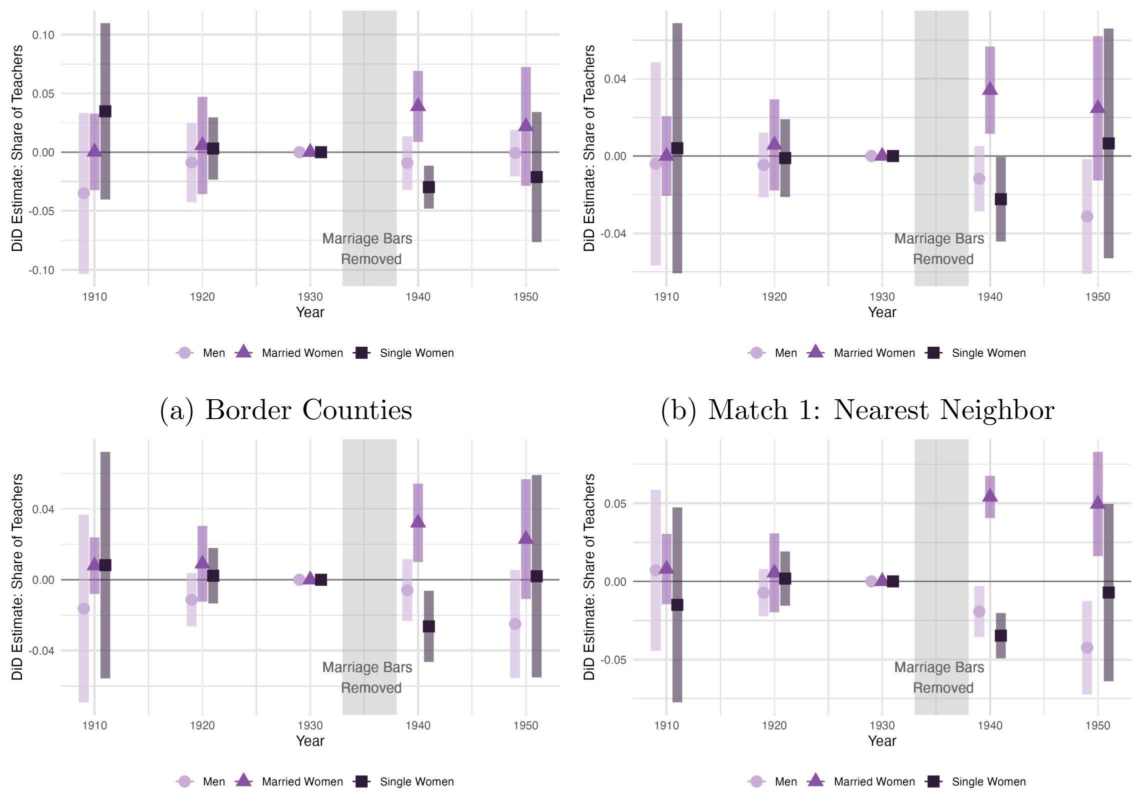
We re-estimate our key analyses using the border county and three matched samples and present the results in Figure C3 and Table C1.

For the border counties and for matched samples 1 and 2, in panels (a) and (b) of Figure C3, our estimated coefficients in 1940 are consistent with our main results—the marriage bar prohibitions caused an increase in the share of married women teachers, at the expense of a decrease in the share of unmarried women teachers, with no change in the share of men in teaching—and significant at the 99% level. Matched sample 3, in panel (d), shows similar results, but suggests a decrease in the share of men in teaching and no convergence by 1950. The slight difference in results for matched sample 3 is consistent with the fact that, as shown in Figure C2, the control group for matched sample 3 is also the least comparable out of all the alternate control groups to the treatment group.

Similarly to our main results, we find no significant evidence of pre-trends in the share of married women, unmarried women, or men teachers prior to 1930 under any of the alternate control groups.

Table C1 uses our linked data to replicate some of our findings on the mechanisms driving this increase in married women's participation in teaching for our alternate control groups. In particular, for matched samples 1 and 2 (shown in Columns (3) and (4)) our point estimates are nearly identical to our main results (replicated in Column (1)), suggesting that the increase in married women teachers was driven by both incumbent unmarried teachers (Panel 1) and women entering the labor force to become teachers (Panels 2 and 3). Matched sample 3 (Column (5)) suggests similar patterns but with larger point estimates. When restricting our sample to only border counties (Column (2)), however, while the point estimates of the effect of marriage bar prohibitions on the propensity of married and unmarried women who were previously not in the labor force (Panels 2 and 3) to get/remain married and become teachers remain similar, the effect of marriage bar prohibitions on the propensity of incumbent unmarried teachers to become married teachers shrinks and while still positive, is no longer statistically significant. This suggests that when focusing only on border counties, the extensive margin channel (of women entering teaching from outside the labor force) was a more important mechanism.

Together, we take this as suggestive evidence that although the marriage bar prohibitions affected the gender composition of teachers in *all* treated states, the mechanisms by which married women entered teaching differed by region.



(a) Border Counties

(b) Match 1: Nearest Neighbor

(c) Match 2: Genetic

(d) Match 3: Full Optimal

Figure C3: Estimated effects of the prohibition of marriage bars in teaching on the gender composition of teachers, at the county level. Estimates are from a difference-in-differences specification where the dependent variable is the share of teachers in a county that are married women, unmarried women, and men. The sample for (a) includes border counties in KY, NC and adjacent states while the sample for (b)-(d) includes all counties in KY and NC and matched control counties, as determined by various methods. Standard errors are clustered at the county level. 95% confidence intervals are shown.

Table C1: Estimated effects of the prohibitions on individual women's propensity be married teachers ten years later.

	Dependent Variable: P(Married Teacher in t)				
Group:	Neighboring (Preferred Spec.)	Border Counties	Match 1: Nearest Neighbor	Match 2: Genetic	Match 3: Full Optimal
	(1)	(2)	(3)	(4)	(5)
Sample 1: Women who were unmarried and teaching in $t-10$					
Treated × 1940 (γ_{1940}^{DD})	0.0217***	0.0131	0.0235***	0.0213***	0.0328***
	(0.0058)	(0.0108)	(0.0060)	(0.0065)	(0.0046)
Dep. Var. 1930 Treated Mean	0.0453	0.0427	0.0450	0.0450	0.0450
Observations	59,542	21,768	62,189	50,638	638,701
Adjusted R^2	0.01861	0.02026	0.01730	0.01472	0.02099
Sample 2: Women who were married and not in the labor force in $t-10$					
Treated × 1940 (γ_{1940}^{DD})	0.0006***	0.0007***	0.0004***	0.0004**	0.0008***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
Dep. Var. 1930 Treated Mean	0.0018	0.0013	0.0018	0.0018	0.0018
Observations	3,125,563	1,290,191	3,230,034	2,758,664	30,967,491
Adjusted R^2	0.00058	0.00060	0.00046	0.00044	0.00096
Sample 3: Women who were unmarried and not in the labor force in $t-10$					
Treated × 1940 (γ_{1940}^{DD})	0.0013***	0.0017***	0.0014***	0.0013***	0.0020***
	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Dep. Var. 1930 Treated Mean	0.0044	0.0040	0.0044	0.0044	0.0044
Observations	2,217,852	858,646	2,146,650	1,902,592	17,910,354
Adjusted R^2	0.00093	0.00120	0.00103	0.00082	0.00163

Notes: Estimation follows Equation (2). For all regressions, the outcome variable is an indicator for whether an individual is both married and a teacher in year t. Column (1) uses individuals in neighboring Southern states as the control group (our preferred specification) and is identical to Column (2) in Table 3. The sample for Column (2) includes border counties in KY/NC and adjacent states, while the samples for Columns (3)-(5) include all counties in KY/NC and matched control counties, as determined by various methods. All regressions included county fixed effects and are clustered at the county level. See Table notes for Table 3 and the Data Section of the main text for further details and full citations for data.

D State-Level Synthetic Difference-in-Differences

Our preferred specification involves analysis at the county level as there is reason to believe that counties with different initial norms and policies surrounding married women teachers may have responded heterogeneously to the uniform 'treatment' of the state-level marriage bar prohibitions. That said, given that the treatment of interest was implemented at the state level, in this section we conduct an alternate state-level analysis to confirm that our results are not driven by our choice of the unit of analysis.

Empirical Strategy

At the state level, we have only two treated units—Kentucky and North Carolina—which motivates our decision to use a synthetic difference-in-differences empirical strategy. We follow Arkhangelsky et al. (2021) in our implementation of synthetic difference-in-differences by finding both unit weights $\hat{\omega}^{sdid}$ that balance pre-treatment trends in our outcome variables for treated and control units as well as time weights $\hat{\lambda}^{sdid}$ that balance trends in control unit outcomes in the pre-treatment and post-treatment periods.[3] We then use the computed weights to solve the following problem:

$$\left(\hat{\gamma}^{sdid}, \hat{\alpha}, \hat{\beta}\right) = \arg\min_{\gamma,\alpha,\beta} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{st} - \alpha_t^{sdid} - \beta_s^{sdid} - \gamma_{1940}^{sdid} \times Treat_s \times Year_{t=1940})^2 \hat{\omega}_i^{sdid} \hat{\lambda}_t^{sdid} \right\}, \quad (1)$$

which boils down to estimating a weighted two-way fixed effects regression to obtain an estimate of our coefficient of interest γ_{1940}^{sdid}, the effect of the prohibition of marriage bars on state-level outcome y_{st} in treated relative to control states. We define our state-level outcomes analogously to the county-level outcomes described in the Outcome Variables Section of the main text.

[3] Because it is unclear which units are treated in 1950, all estimation in this section uses only data from 1910-1940.

Estimation and Results

We begin by computing the unit and time weights. State weights $\hat{\omega}^{sdid}$ as computed for our main outcome variable (the share of women in state s who are married women) are shown in Panel (a) of Figure D1. Computed time weights for all analyses are 1 for 1930 and 0 for other pre-period years, simplifying our setup to a two-period setting.

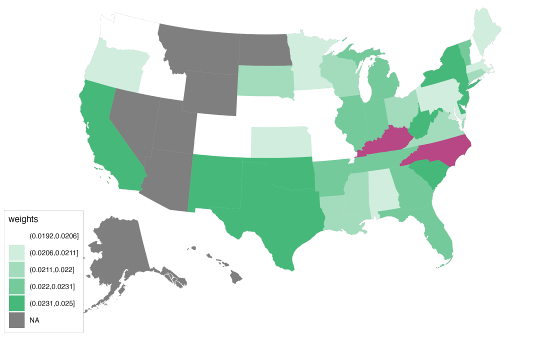
First, we use the synthetic DiD setup to estimate the state-level effect of the marriage bar prohibitions on the composition of the teacher workforce. Figure D2 plots estimates of γ_{1940}^{sdid} for the share of teachers who were married women (triangles), men (circles), and unmarried women (squares) respectively. The first two groups of estimates juxtapose the results our preferred specification (at the county level with county-level clustered standard errors) with the state-level synthetic DiD estimates. Standard errors are computed using a permutation approach ('placebo' option in R). The point estimates are very similar between the county and state-level analyses, and our primary finding that the share of married women teachers increased in treated relative to control states remains significant at the 95% confidence level despite widened confidence intervals.

An additional benefit of using the synthetic DiD approach is that it also allows us to perform analysis with only one treated unit. As such, we estimate the effects of the prohibitions on married women teachers in North Carolina and Kentucky separately. Panels (b) and (c) of Figure D2 map control unit weights for NC and KY individually, and we see that much of the control unit weighting in the joint analysis is driven by North Carolina. Nevertheless, the third and fourth groups of estimates in Figure D2 show that, while noisier, the point estimates for the effects of the marriage bar prohibitions on NC and KY are remarkably similar to the overall effects, suggesting that results are not being driven by one state alone.

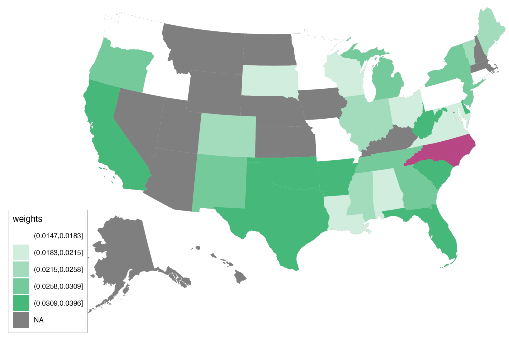
Finally, we replicate our results on mechanisms using the state-level synthetic DiD and linked data aggregated to the state level. Table D1 replicates Table 3 from the body of the paper, presenting synthetic DiD estimates and standard errors computed using the placebo method. Note that this *not* an individual level regression and therefore not directly comparable to Table 3 since it is weighted to match pre-trends rather than by population.

Results are qualitatively similar to our preferred estimates. We therefore conclude that our main results are not sensitive to the chosen unit of analysis.

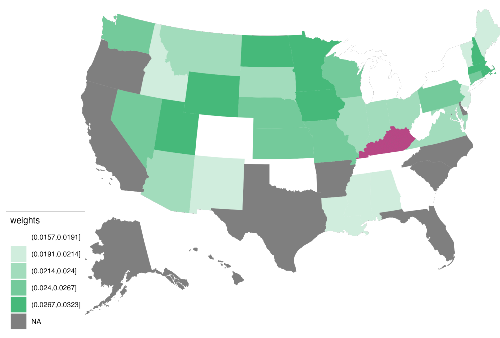
Figure D1: Maps of treated (pink) and control (green) states shaded by unit weights computed in our synthetic difference-in-differences empirical strategy (computed to match pre-trends between control and treatment states for outcome variable share teachers married women). Panel (a) uses both Kentucky and North Carolina as treated units, panel (b) uses only North Carolina as a treated unit, and panel (c) uses only Kentucky as a treated unit.



(a) Control Unit Weights: North Carolina and Kentucky



(b) Control Unit Weights: North Carolina Only



(c) Control Unit Weights: Kentucky Only

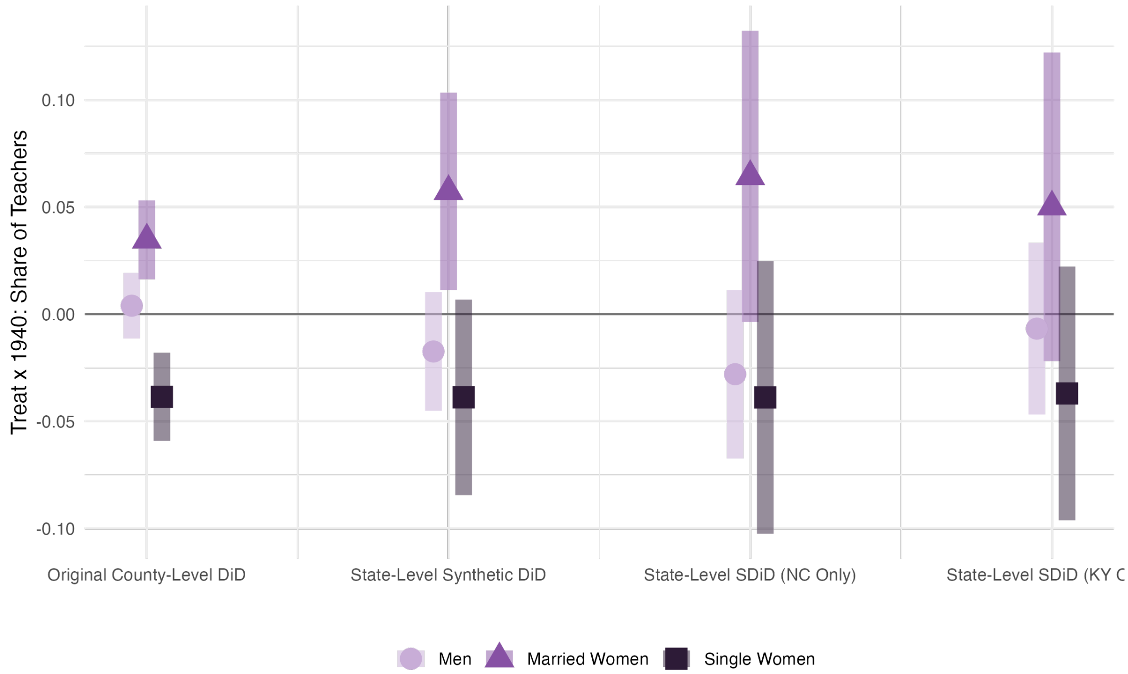


Figure D2: Estimates of the effect of the prohibition of marriage bars in teaching on the gender composition of teachers. The first column uses the standard difference-in-differences setup from our main specification at the county level, with standard errors clustered at the county level. The other columns use state-level synthetic difference-in-differences, with standard errors computed using a 'placebo' method. The second column includes both KY and NC as treated units, while the third and fourth only include NC and KY respectively. 95% confidence intervals are shown.

Table D1: State-Level Synthetic DiD: Estimated effects of the prohibitions on women's propensity to get married and either teach, work outside of teaching, and exit the labor force.

Dependent Variable:	Pr(Married in t)	Pr(Married Teacher in t)	Pr(Married Non-Teacher in LF in t)	Pr(Married Not in LF in t)
	(1)	(2)	(3)	(4)
Sample 1: Women who were unmarried and teaching in $t-10$				
Treated × 1940 (γ_{1940}^{DD})	-0.033	0.036^{***}	-0.003	-0.082^{**}
	(0.035)	(0.013)	(0.006)	(0.038)
Dep. Var. 1930 Mean	0.5857	0.0452	0.0350	0.5056
Sample 2: Women who were married and not in the labor force in $t-10$				
Treated × 1940 (γ_{1940}^{DD})	-0.003	0.0007	0.009	-0.013
	(0.002)	(0.0005)	(0.009)	(0.009)
Dep. Var. 1930 Mean	0.9328	0.0018	0.0549	0.8761
Sample 3: Women who were unmarried and not in the labor force in $t-10$				
Treated × 1940 (γ_{1940}^{DD})	-0.024	0.002^{**}	0.011^{***}	-0.038^{*}
	(0.017)	(0.0008)	(0.004)	(0.021)
Dep. Var. 1930 Mean	0.5072	0.0044	0.0437	0.4590

Notes: Estimation follows Equation (1). Construction of these state-level linked samples follows the construction of individual-level linked samples, but aggregated to the state level and weighted according to the procedure outlined above. Sample 1 contains linked women who were under 40, unmarried, and teaching in 1910, 1920, and 1930, Sample 2 contains linked women who were aged 18-50, married, and not in the labor force in 1910, 1920, and 1930, and Sample 3 contains linked women who were aged 8-40, unmarried, and not in the labor force in 1910, 1920, and 1930. All regressions use the 1910-1920, 1920-1930, and 1930-1940 linked full-count Census samples. Standard errors are computed using a 'placebo' method. See the Data Section of the main text for details and full citations for data.

E Linkage Rates

This appendix describes the linkage rates between consecutive censuses and discusses potential concerns about the bias that the linking process may introduce to the results.

Linkage rates. Online Appendix Table B1 shows linkage rates for various populations across censuses. While linkage rates are largely similar over time and between treated and control states, there are two differences of note. First, linkage rates are higher in all years for married women versus unmarried women (65.8% versus 53.9% in 1920) and for white women versus Black women (62.0% versus 32.1% in 1920). These differences are known in the literature, in part due to the Census Tree links coming from a free genealogical website (Buckles et al. 2023), and persist over time. As a result, women who get married or move between decennial Censuses are less likely to appear in our linked samples, which may attenuate our estimated effects of the prohibitions on unmarried women teachers towards zero.

Second, linkage rates increase over time for married women but remain stable for unmarried women. One potential explanation for these trends could be that reports on the FamilySearch website are more frequent in more recent years, driven by descendants of married women. However, given that linkage rates rise similarly in both treated and control states, we are not concerned that the differential trends bias our results.

Bias due to linkage loss. Linkage loss could introduce bias to our results if linkage rates differed by women's treatment status, marital status, or employment status. We find similar linkage rates between treated and control states, so the first case is not a concern. However, the second and third cases may both be at play. On one hand, because single women who get married between censuses are less likely to be linked, single women who get married are underrepresented in our linked samples. As a result, using the linked data, our estimated effects of the prohibitions on the likelihood that a woman in $t-10$ is married in t are attenuated towards zero. On the other hand, single women who stop working because they got married between censuses are also underrepresented in our linked samples. As a result, using the linked data, our estimated effects of the prohibitions on the likelihood that a woman in $t - 10$ is working in t are upward biased.

It is difficult to bound these biases, as doing so would require estimating the number of single women who get married whom we are unable to link, which is inherently unobserved. However, what we can say is that our main estimate of interest in our linked analyses—that is, the effect of the prohibitions on the likelihood that women in $t - 10$ become married women teachers in t—is ambiguously biased due to linkage loss, given it is subject to two sources of bias that go in opposite directions and could even negate each other.

References

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. "Synthetic Difference-in-Differences." *American Economic Review* 111, no. 12 (2021): 4088–4118. https://doi.org/10.1257/aer.20190159. https://www.aeaweb.org/articles?id=10.1257/aer.20190159.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley Wilbert. *Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project*, NBER Working Paper No. 31671. September 2023.

Diamond, Alexis, and Jasjeet S Sekhon. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *The Review of Economics and Statistics* 95, no. 3 (2013): 932–945. Accessed February 7, 2024.

Fischer, Alexander, and David Roodman. *fwildclusterboot: Fast Wild Cluster Bootstrap Inference for Linear Regression Models (Version 0.14.3)*, 2021. https://cran.r-project.org/package=fwildclusterboot.

Fishback, Price V, William Horrace, and Shawn Kantor. "Did New Deal Grant Programs Stimulate Local Economies? A Study of Federal Grants and Retail Sales during the Great Depression." *The Journal of Economic History* 65, no. 1 (2005): 36–71.

Hansen, Ben B. "Full Matching in an Observational Study of Coaching for the SAT." *Journal of the American Statistical Association* 99, no. 467 (2004): 609–618.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth A. Stuart. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42, no. 8 (2011): 1–28.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. *The Census Tree, 1910-1920.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], September 2023.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. *The Census Tree, 1920-1930*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], September 2023.

———. *The Census Tree, 1930-1940*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], September 2023.

Ruggles, Steven, Matt A Nelson, Matthew Sobek, Catherine A Fitch, Ronald Goeken, J David Hacker, Evan Roberts, and J Robert Warren. *IPUMS Ancestry Full Count Data: Version 4.0 [dataset]*. Minneapolis, MN: IPUMS, 2024.

Sekhon, Jasjeet S. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software* 42, no. 7 (2011): 1–52.