

RE: The Journal of Economic History - Decision on Manuscript ID JEH-2024-0189

Dear Referee #1,

Thank you for your helpful report. We have done our best to reply to each of your concerns, which we feel has dramatically improved the paper.

Before delving into our point-by-point response to your letter, we wanted to summarize some of the major changes to the draft.

First, we have expanded our discussion of the historical context considerably. Our identification strategy, results, and robustness sections now include careful discussions of how the Great Depression and New Deal may have interacted with the prohibitions we study and the effects we find.

Second, we have updated all our analysis using the linked Census data to use individual-level data, as opposed to county-level aggregates.

In what follows, we respond point-by-point to questions and concerns raised in your report (original comments are in *blue italics*, our replies in regular font).

Thank you again.

## Referee #1's Comments

*Major: I have no major concerns.*

### Authors' reply:

Thank you!

*Minor: The biggest minor concern that I have is that I'm not sure that 1910 should be in your main specifications. It makes more sense to have two decades before if you're having two decades after. (Or it might make sense just to compare 1930 to 1940, since 1950 isn't in the linked data.) I don't think removing 1910 will affect your results given Figure 2 (and it should still be shown in Figure 2). Similarly in the linked data, I don't understand why you have 1910-1920 when you only have 1930-1940 in the post. (And then Table 5 doesn't use 1910-1920?)*

### Authors' reply:

Thank you for this suggestion, and also for catching the typo in the notes of Table 5: we do use 1910-1920 in Table 5 as well, and have corrected the notes accordingly.

It is true that removing 1910 does not affect our main results. Our main reason for including 1910 is to explore the possibility of pre-trends in our main specification, as shown in our Figure 2 (replicated below for your convenience). We do construct a balanced panel of counties from 1910 through to 1950, so although we are comparing 3 decades prior to 1930 to 2 decades after, we believe that balance should not present a problem in our main specification. However, we also verify that the results are robust to including only one decade before and after the policy changes our results.

*For footnote 1, I don't think that this is necessarily a good reason to omit non-White women (1. It isn't teacher specific, 2. More detail is needed about why the laws wouldn't apply to married Black women—do Black school districts just not use them?). I think instead you should elevate A4 and A5 to be part of Figure 2. The odd pre-trends for Black women seem reason enough, and the fact that putting them in with the White women doesn't affect the results (presumably the sample is smaller) helps mitigate concerns about p-hacking. I would also find some historical evidence that Black and White teachers are not substitutes in the Jim Crow South so the patterns for Black teachers are unlikely to be affecting patterns for White teachers as they are different schools. (Maybe Carruthers and Wanamaker have a paper?)*

### Authors' reply:

Thank you for this advice. Following your suggestion, we have elevated the former appendix figures showing the results for county shares of Black teachers to the main text, and now

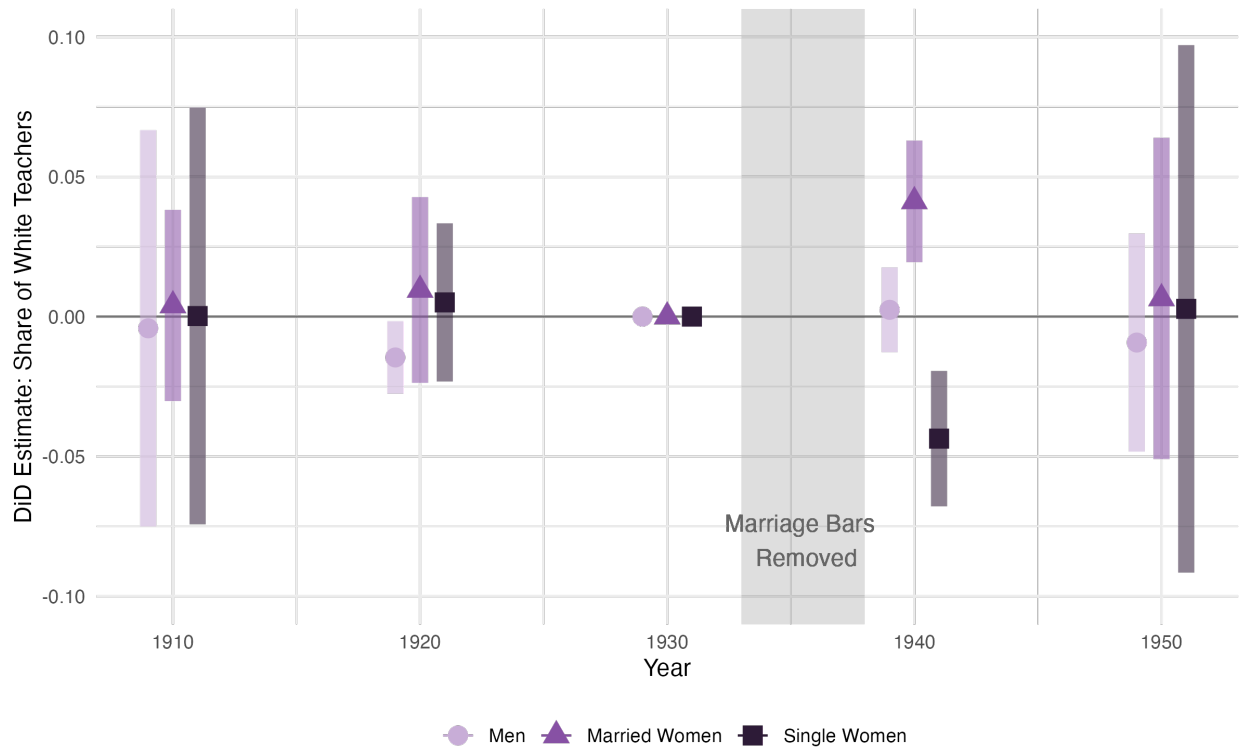
discuss the results first for the share of *all* teachers who are men/unmarried women/married women, and then separately for white and Black women. Figure 2 now shows the overall effect of the prohibitions for the teaching population, while Figure 3 shows the results separately by race. We replicate both figures below for your convenience.

Because schools were segregated during our time period of study, Black and white teachers were indeed not substitutes under Jim Crow era laws. We include mention of this fact when discussing the results in Section 5, reproduced here:

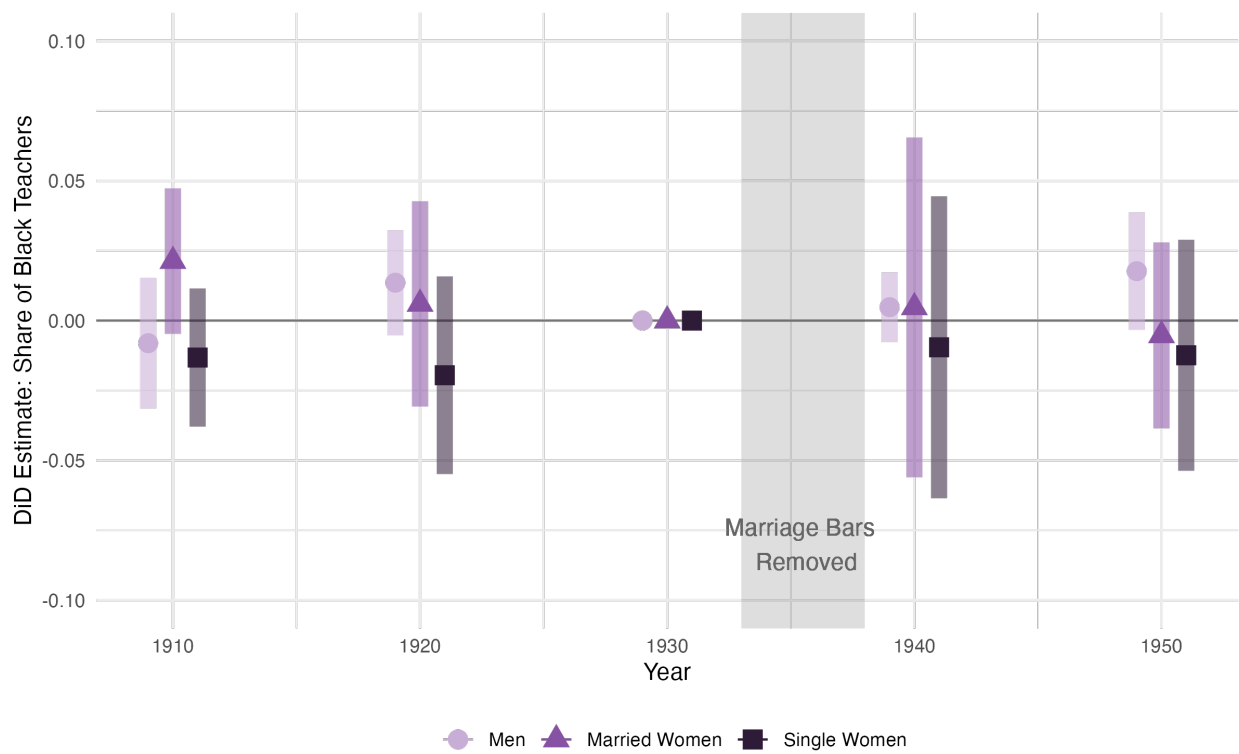
We interpret these findings as evidence that marriage bars were binding for white women but not Black women, as posited by [11]. Our findings are consistent with the fact that white and Black women in this setting worked in segregated schools due to Jim Crow era laws, and the fact that Black married women were more likely to work during this time period in general ([6]).



**Figure 2.** Estimated effects of the prohibition of marriage bars in teaching on the *gender composition* of all teachers, at the county level. Sample includes KY, NC, and neighboring Southern states.



(a) Effects on composition of white teachers.



(b) Effects on composition of Black teachers.

**Figure 3.** Estimated effects of the prohibition of marriage bars in teaching on the gender composition of (a) white and (b) Black teachers at the county level.

*In the background, provide a little information on how difficult it was to get a teaching degree. How long does it take to complete normal school (do they need to complete normal school)? Does this change during your time period in your states? (The historical examples I know take 1-3 years, but are not in the South.) (It might be helpful to have a section in the lit review on the teaching environment—what did these schools look like? Were there one-room school houses? Were there separate grades? Did this change over time? Were schools segregated? Etc.)*

**Authors' reply:**

Thank you for this suggestion. Because what matters most for our design is whether there was variation in teacher training across states and over time, for the sake of brevity, we have added a mention of the most important change in teacher training during 1910-1950.

The primary change that occurred in teacher training, which we now mention in the identification strategy in Section 4.3, took place before 1930 in both our treated and control states:

One of the key changes in teacher training in the early 1900s was the shift from 2-year “normal schools” to 4-year college programs, but these changes largely took place before 1930.

From this, we do not expect changes in teacher training to affect our analysis.

*Think about other things that might be going on in your treatment states compared to your control states that could affect married vs. single women in the labor force and argue or demonstrate that they are not problematic. What have you come across historically? (Ex. did time to degree change differentially? Did marital property laws change differentially? Child labor laws?—this is probably irrelevant because you find no effects on fertility. Compulsory schooling laws? Any depression-era projects? I’m not sure what else might be relevant.) I think it is unlikely that there’s anything else relevant in the 1930s that just affects your two treatment states or affects your control but not treatment states, but it would be helpful to double check and then state what you checked.*

**Authors' reply:**

Thank you for this suggestion. We agree that given the timing of the prohibitions, we certainly should provide more contextual details as to how the Great Depression, New Deal, and other policies may have affected married and single women and thus interact with our findings.

In the updated text, we investigate how various economic and policy changes differentially affected our treated and control states, assess how our results may be biased as a result, and include robustness checks that allow us to assess the extent of the bias empirically.

The main new text can be found in Section 4.3. For your convenience, we replicate the new text below:

Despite the many economic and policy changes that took place between 1910 and 1950 that could have affected women’s marriage and employment outcomes, many of these changes did not differentially impact our treated and control states. For instance, changes in compulsory schooling laws and child labor laws, which could have affected demand for teachers, did not occur in most of our treated and control states between 1910 and 1950.<sup>7</sup> One of the key changes in teacher training in the early 1900s was the shift from 2-year “normal schools” to 4-year college programs, but these changes largely took place before 1930.

However, we do find two notable differences in how our treated and control states were affected by the Great Depression and New Deal. Here, we reference prior work on the effects of the Great Depression to assess the potential biases these differences may cause in our design, and develop strategies that allow us to address these concerns.

The first difference is that the average treated county experienced a more severe economic downturn and a more muted recovery than the average control county during the Great Depression.<sup>8</sup> Two control states in particular experienced comparatively mild effects of the Great Depression: in Supplemental Table B2, Column (2) shows that while KY/NC experienced similar decreases in retail sales per capita between 1929 and 1933 as control states TN and WV, SC and VA experienced much smaller declines. At the same time, Column (3) shows that with the exception of WV, control states experienced slightly larger recoveries in retail sales per capita between 1933 and 1939 than KY/NC.

How might these differential trends bias our results? Prior literature finds that the Great Depression delayed women’s decisions to marry ([17]) and contributed to more widespread implementation of marriage bars, to preserve job openings for men ([11]). At the same time, [4] find that more severe economic downturns pushed more white women into the labor market, with suggestive evidence that married women were affected as well—although this “added worker” effect was tempered by New Deal programs ([8]).<sup>9</sup>

The direction of potential bias is therefore ambiguous. In the absence of the marriage bar prohibitions, worse economic conditions in treated states relative to control states would have increased the use of marriage bars, delayed marriage among women, and decreased resources for schools, all of which would have reduced the relative likelihood that women got married or taught in treated

---

<sup>7</sup>The only change that occurred was that VA, one of our control states, required one more year of compulsory schooling in the 1920s, which would have led to increased demand for teachers and bias our estimates downwards.

<sup>8</sup>There is data on New Deal spending by county from 1933 to 1939, but given the time period and decennial nature of the census, the data cannot be used to construct time-varying measures of New Deal spending in 1930 and 1940. We proxy for both recession severity and economic recovery using changes in retail sales per capita instead.

<sup>9</sup>Similarly, while prior work finds that birthrates decreased during the Great Depression ([23]), the New Deal countered these effects by decreasing infant mortality and increasing birthrates ([9]).

states compared to control ones. At the same time, the “added worker” effect may have resulted larger inflows of married women into the labor force in treated states compared to control states. Our estimates could thus be biased downward by the reduced likelihood of treated women getting married or teaching in treated states, or biased upward by the “added worker” effect in treated states.

We use several strategies to address these potential biases. Our primary strategy involves comparing treated states to alternate control groups that were more similarly affected by the Great Depression. Supplemental Appendix C outlines these specifications in detail, including a border counties design, where we only compare counties along state borders, and a matched county design, where we match counties based on 1920 and 1930 characteristics. We also include controls for unemployment rates and industry shares in several specifications.

The second notable difference between treated and control states was their exposure to New Deal policies. A large number of control counties in TN, but only a small number of treated counties, were directly impacted by the Tennessee Valley Authority (TVA), a major New Deal program that aimed to provide the first federal electrification program and stimulate the economy in the hard-hit Tennessee Valley. Prior work finds that the TVA had large impacts on the local economy, particularly in shifting work away from agriculture to higher-paying manufacturing jobs ([18]). Combined with the aforementioned studies, these results suggest that including TVA-affected counties in our analysis could bias our estimates upwards, as increasing manufacturing jobs in TVA-affected counties would have decreased the need for married women to work in more control counties than treated ones. To address this potential bias, we conduct robustness checks that exclude the counties affected by the TVA.

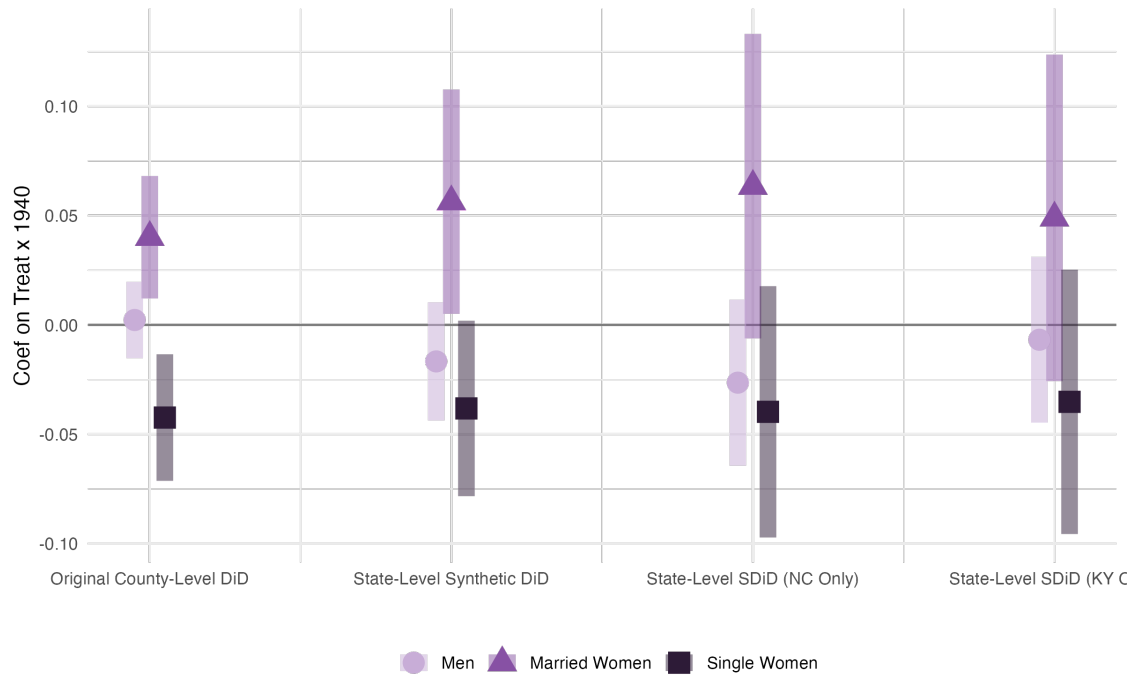
We describe the results of the additional specifications we include to account for the effects of the Great Depression and New Deal in Section 5:

Finally, we discuss the robustness of our results to the impacts of the Great Depression and New Deal. Our alternate control group designs allow us to address the potential bias caused by treated states being more severely impacted overall by the Great Depression than control states. In our border counties design, we compare counties that share similar industry composition and other unobservable characteristics before 1933, and hence are more likely to experience similar effects of the Great Depression and more similar New Deal support. In our matched counties design, we match counties on retail sales per capita in 1929 and growth over time. We also examine how our results are affected by including controls for county unemployment rates and industry composition. Finally, in Column (5) of Supplemental Table B6, we present our results excluding counties in the TVA. As discussed above, our results are robust to all these specifications, suggesting that any biases caused by the Great Depression are minimal.

*Are your results different for KY and NC?*

### Authors' reply:

This is an important question, especially since the state prohibitions were implemented in different years. To explore potential heterogeneity across states, Supplemental Appendix Figure D2 shows the main difference-in-difference estimates separately by state. We reproduce the figure below for your convenience.



**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.

The results are qualitatively the same by state, although the estimates are less precise given the sample becomes quite small. In line with what one might expect, the effect is slightly weaker in KY, where the laws were not in effect as long as they were in NC by 1940.

We also provide a footnote in the empirical strategy section that mentions this difference between states:

In our state-level analysis in Supplemental Appendix D, we also show results for NC and KY separately. Our findings are not being disproportionately driven by one state; in fact, point estimates are remarkably similar for both states.

*Footnote 4: Be more clear about what you do when you say you condition on women's employment and marital history. When I first read this comment, my thought was, isn't that*



*endogenous to the laws? But I think you mean you're using a matched sample following the same women to condition on their employment and marital history prior to the laws, which would not be endogenous.*

**Authors' reply:**

Thank you for noting this. We have since removed this footnote for brevity, but as we discuss in a later response, we have clarified our usage of the word “conditional” throughout the paper.

*Do the laws affect educational attainment of affected women at all (that is, are girls in high school(?) more likely to invest in education if they know they can be employed long-term as a teacher, whether or not they actually become a teacher?) This would be separate from the question of if teacher retention affects quality.*

*Speaking of quality—20th century literature suggests that there are large returns to experience for the first few years of teaching. Could students have better outcomes with less teacher turnover as a result of this law or do your results preclude that possibility?*

**Authors' reply:**

Thank you for these suggestions. We had not looked at the effects on student outcomes and were very interested to see what we would find. Unfortunately, the data on student outcomes in the early 20th century are extremely limited. Without standardized testing or data on educational attainment (which was only collected in the Census beginning in 1940), our only option would be to look at school attendance as recorded in the Census over our time period of interest. However, there were a few changes in how school attendance was collected in the Census over time which preclude us from conducting a reliable investigation here.

The main issue is that the Census question on school attendance changed from asking about attendance/enrollment “in school in the last six months” in 1930, to attendance/enrollment “in the last month” in 1940, and then again to attendance/enrollment “in the last two months” in 1950. Since the Census was always conducted on April 1st, this means that the school attendance question covers September to April in 1930, March to April in 1940, and February to April in 1950. If we ignore this change in the wording and naively plot “attendance” across the decades, we find a general upward trend in “attendance” across all states from 1910 to 1950, but an unusually large drop in attendance in Kentucky in 1940, which appears to recover back to a similar level as the neighboring states in 1950.

The size of the drop would be concerning for our parallel trends assumption in our main analysis if it were “real”—but it is unclear that it is. Historical records on attendance statistics from reports in Kentucky in the 1930s and 1940s do not show a large decrease in attendance in 1940. Instead, the main explanation we find for the drop in the Census in 1940 is that the Kentucky school term was likely shorter than other states: in 1930-31, Kentucky's average school term was 159 days, compared to an average of over 170 days in the neighboring states [10]. So, if Kentucky schools mostly ended in late February, the 1940 version of the school

attendance question would miss many students who did in fact attend school in the 1939-40 school year, explaining why attendance appears to drop suddenly in 1940 in Kentucky but recover by 1950. Because these school term lengths did not change between 1930 and 1940, we are not concerned that these differences pose a problem for our parallel trends assumption; however, due to this issue, we do not incorporate results on the effects of the prohibitions on educational attainment of women or children.

*p.11 What kinds of bans did KY and NC have before? Do you have any (suggestive) evidence? Do you have any way of knowing if the effects are different by type of ban (maybe at the county level?) or would that be too much/impossible data collection?*

**Authors' reply:**

Thank you for this question—it is one that we would love to be able to answer. Unfortunately however, we are not aware of county-level data on marriage bars prior to the passage of the prohibitions we study. There are some journalistic pieces that describe the use of marriage bars within particular districts: for instance, some of which were published in the newspapers we cite when describing the marriage bars in the context section, but these serve more as case studies. As such, we take a “data-driven” approach in the paper and assess the extent to which marriage bars were in place by looking at how the shares of married women in teaching evolved over time across counties.

*I had a difficult time understanding why many of the results were attributed to hiring of married women as opposed to retention of married women. On page 22, you find that retention is important. Similarly, are you sure that unmarried teachers are being “let go” (p.24) as opposed to not hired?*

**Authors' reply:**

Thank you for raising these questions. We have clarified the language throughout our results section to better distinguish between the extent to which our results indicate changes in retention and hiring practices. To summarize: our linked analysis allows us to distinguish between retention and hiring channels in two steps. First, we isolate the group of women who could *only* have contributed to the increase in married women teachers through retention: that is, the women who were previously unmarried teachers. For these women, we estimate the effects of the prohibitions on the likelihood they became married teachers. We then do the same for the group of women who could *only* have contributed to the increase in married women teachers through hiring: that is, the women who were already married but not in the labor force. Comparing results between the two groups allow us to assess the role of retention and hiring practices in explaining our main results.

We quote the results of this analysis below:

Our results suggest that the increase in married women in teaching was driven entirely by changes in extensive margin labor supply, both by increasing *retention*

of incumbent teachers who would otherwise have exited the labor force, and by increasing *hiring* of women who would have otherwise remained out of the labor force. We find no evidence of negative effects on the LFP of women who were already in the labor force or who would have entered the labor force even absent the prohibitions.<sup>10</sup>

Furthermore, our estimates suggest that new hiring contributed more to the increase in married women teachers than the retention of incumbent teachers did. To see this, we scale the estimated effects in Table 3 by the total number of treated white women in each group in 1930. Although the effect size in percentage points is largest for unmarried women teachers (2.2 p.p.), there were fewer than 8,000 such women in our treated linked sample in 1930, suggesting the prohibitions led to the retention of an additional 168 married women teachers. In contrast, there were more than 434,000 married women outside of the labor force in our treated linked sample in 1930, meaning our effect size of 0.06 p.p. translates to an increase of 258 married women teachers due to new hiring. We conclude that roughly 60% of the overall increase in married women teachers was due to the hiring of new married women.<sup>11</sup>

*On page 13 where you say small numbers of Black teachers, provide the actual numbers.*

**Authors' reply:**

Thank you for this suggestion. We have since removed the paragraph you refer to—in which we originally motivated why our analysis focuses on white teachers only—as we now present our main results separately by Black and white teachers, following your earlier suggestion.

*p.14 top: Linkage rate problems—do you have any thoughts on the potential direction of biases?*

**Authors' reply:**

Thank you for raising this question. We have added a discussion of the potential biases that may arise from linkage issues to the data section, and replicate the new text below for your convenience:

---

<sup>10</sup>See Supplemental Table B4 for the estimated null effects on married and unmarried non-teachers in the labor force in  $t - 10$ .

<sup>11</sup>While women who were unmarried and outside the labor force in  $t - 10$  also contributed to the increase, recall that we cannot disentangle what shares of these women became married women teachers through hiring or retention channels. However, we can apply a similar argument: of the unmarried women out of the labor force in  $t - 10$ , a much larger number will get married by  $t$  than become teachers by  $t$ ; hence, if the percentage point effect sizes are similar in magnitude, then the effect of the prohibitions on unmarried women outside of labor force will likely be driven by hiring effects.

Supplemental 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 ([5]), 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.

*When you discuss Table 1, I think you're trying to do two things: 1. Show how well the southern states you've chosen work as a control group for KY/NC, and 2. What is the external validity of your results to other states. I think making the motivation more clear up front (why are you providing this information) would help when you discuss the summary statistics. Also, I would start with discussing your preferred comparison group rather than ending with it (since economists generally care about internal validity first and external second).*

### Authors' reply:

Thank you for this advice. We have updated the discussion of Table 1 such that it is both more concise and immediately addresses the main goal of the table, which as you pointed out, is to demonstrate the internal validity of our design. We replicate the new text below for your convenience:

Our analysis focuses on the counties within and surrounding the states that passed prohibitions (KY/NC). To motivate this focus, Table 1 provides summary statistics comparing counties in KY/NC to counties in other states in 1930.

KY/NC were highly comparable with their neighboring Southern states, SC, VA, TN, and WV, as shown in Columns (3) and (4). Compared to the national average (Column (1)), counties in KY/NC and neighboring states had higher student-to-teacher ratios, exhibited a lower share of the population living in urban areas, and saw more children per married woman on average.<sup>12</sup> Within the

<sup>12</sup>These statistics also highlight that our findings are most relevant for the U.S. South. The higher student-teacher ratios in particular in Southern states are consistent with anecdotal evidence of teacher shortages in the South, as in [12].

U.S. South (Column (2)), counties in KY/NC and neighboring states tended to have lower rates of LFP among white married women, particularly in teaching. There are a few small but notable differences between KY/NC and the neighboring states—in 1930, neighboring states were more urban and had higher LFP among married women, higher unemployment rates, and greater shares of single women in teaching—which warrant robustness exercises using controls and different empirical strategies.<sup>13</sup> However, given the overall cultural and statistical similarities between KY/NC and their neighboring states, the neighboring states comprise our preferred comparison group.

Note that we also add Footnote 12 within the paragraph that speaks to the external validity of our findings.

*p.15 Explain why you are doing things at the county level.*

**Authors' reply:**

Thank you for raising this point. The main reason we present our main results using county-level outcomes in the cross-sectional analysis because with this analysis, we are interested in assessing the effects of the prohibitions on a district's teacher workforce as a whole, meaning our outcomes of interest are most naturally defined at a regional level.

However, as the editor and Referee #2 point out, there is more precision and information offered in the individual-level data which can be leveraged in our linked analysis in particular. As a result, we have updated all of our linked analysis to use individual-level data. We also replicate our main county-level cross-sectional results at the individual level in Supplemental Table B3, and Column (1) shows we find effects of similar size and magnitude as in our county-level specification. We provide the table below for your convenience.

---

<sup>13</sup>See Supplemental Appendix C.

**Table B3.** Heterogeneity in estimated effects of prohibitions on becoming a married woman teacher

Dependent Variable: Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Variables</i>									
Treated $\times$ 1940 ( $\gamma_{1940}^{DD}$ )	0.0350*** (0.0073)	0.0347*** (0.0066)	0.0240*** (0.0079)	0.0313*** (0.0064)	0.0375*** (0.0075)	0.0388*** (0.0072)	0.0346*** (0.0072)	0.0405*** (0.0077)	0.0397*** (0.0068)
Treated $\times$ 1940 $\times$ <i>Nonwhite</i>						-0.0332** (0.0129)		-0.0426** (0.0171)	
Treated $\times$ 1940 $\times$ <i>Urban</i>							-0.0023 (0.0112)	-0.0107 (0.0121)	
Treated $\times$ 1940 $\times$ <i>Nonwhite</i> $\times$ <i>Urban</i>								0.0306 (0.0232)	
Treated $\times$ 1940 $\times$ ( <i>Age</i> - 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 <sup>2</sup>	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 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. Column (1) estimates Equation (2) for 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 Section 3.3 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 Section 4.3 for further discussion). Columns (6) and (7) include indicator for whether a teacher is non-white (*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 ([22]). Standard errors are clustered at the county level.

*P.16 top: do you mean year or decade?*

**Authors' reply:**

Thank you for this clarifying question. When we wrote that “ $t$  indexes year” in our main diff-in-diff specification, we meant the *census year*, e.g., 1920, 1930, etc. We have replaced the word “year” with “census year” for clarity.

*Good job on explaining why you're doing what you're doing re: clustering. I had the question you answered in the next sentence.*

**Authors' reply:**

Thank you!

*Footnote 23: Dropping: Why? How many do you lose? Is there evidence of selection? (Also are these dropped in your Table 1?)*

**Authors' reply:**

Thank you for raising this question. The original footnote described how we construct our sample of counties for the original linked analysis. However, because we have changed our linked analysis to be at the individual level, obviating the need to drop counties, we have now removed this footnote.

*P. 19: TN, WV—are these not treated by the court decisions (they are in your control group)? When did the court decisions happen?*

**Authors' reply:**

Thank you for raising this point. In the paragraph you refer to, we describe newspaper reports of courts that ruled against school boards' right to dismiss teachers on the basis of their marital status in some of the control states (TN and WV, specifically). However, these were *local* court decisions, rather than state-wide prohibitions, like those we study in KY and NC. If anything, these local prohibitions occurring across both treated and control states bias our main estimates downward, but we do not believe this bias is significant, given these court cases were highly idiosyncratic.

*Put footnote 27 in the paper proper.*

**Authors' reply:**

Thank you for this suggestion. We have now moved the footnote on whether the schools may have responded to the prohibitions by changing the margin on which they discriminated against married women (from discriminating based on marriage to discriminating based on whether women had children) into the main text.

*When you discuss your mechanisms results, be a little more clear and remind people what you're doing for each of these specifications. They're more complicated than the previous results and it is difficult to understand what you're doing without looking at the tables at the same time.*

### Authors' reply:

Thank you for this suggestion; we indeed found that the linked analysis certainly was more complicated to explain than the main cross-sectional results!

We find that the results are more straightforward now that our linked analysis uses individual-level data rather than county-level aggregates. All our estimates in the section can now be interpreted as the estimated effect of the marriage bar prohibitions on the likelihood that a woman in a particular demographic group in  $t - 10$  transitioned into a particular marital status and employment status in year  $t$ . For instance, we find that the prohibitions increased the likelihood that women who were unmarried teachers in  $t - 10$  became married teachers in year  $t$ , and so on.

To reflect this change, we have revised our opening paragraphs to the discussion on the mechanisms section in Section 5 as follows:

What are the mechanisms driving our main effect? Did the prohibitions pull married women into the labor force (an 'extensive margin' effect), or did the prohibitions only affect women who would have worked even absent the policy (an 'intensive margin' effect), by either inducing unmarried teachers to marry or by inducing married workers to switch to teaching?

We answer this question using our linked Census data, which allows us to trace out individual transitions into marriage and employment for three groups of white women, grouped by their marital and employment status in year  $t - 10$ : (1) unmarried women teachers, (2) married women not in the labor force, and (3) unmarried women not in the labor force.<sup>14</sup> Results from estimating Equation (2) for each group are shown in Table 3.

*Do you have any findings on wages?*

---

<sup>14</sup>Given our findings that the marriage bars had no direct effects on Black teachers, we focus on white teachers in our analysis of mechanisms. In addition, in Supplemental Table B4, we examine outcomes for other working women. We restrict our linked samples of unmarried women to be between the ages of 8 and 40 (16-40 for teachers) in  $t - 10$ , and we restrict our linked samples of married women to be between the ages of 18 and 50 in  $t - 10$ , to focus on the populations for whom the decisions to marry and work were the most relevant. Our results are robust to alternate cutoffs.



**Authors' reply:**

Thank you for this question. Unfortunately, since the Census did not collect data on earnings prior to 1940, we are unable to investigate the effects of the prohibitions on earnings. We also cannot use occupational scores, the usual proxy for income [1], because they are occupation-based and would thus not be able to capture any changes in teaching wages over time.

However, because this question is undoubtedly important and of interest in our setting, we do try to gain some traction on the effects on income by looking at the effects of the prohibitions on the occupational scores of incumbent unmarried teachers. The results are shown in Supplemental Table B5, which we replicate below for your convenience.

If incumbent unmarried teachers were pushed out to other lower paying occupations or were mostly pushed to unemployment, we would detect a decrease in their occupational scores in 1940. However, we find no such effect, suggesting that those who switched professions were not pushed to lower paying jobs on average.

**Table B5.** Estimated effects of marriage bar prohibitions on fertility, occupational scores, and mobility

Dependent Variable:	Has child in $t$		Occupational Score in $t$		Moves state in $t$	
Model:	(1)	(2)	(3)	(4)	(5)	
Treated $\times$ 1940	-0.0208 (0.0135)	0.7275* (0.4080)	-0.0949 (0.2835)	0.0401*** (0.0093)	0.0593*** (0.0118)	
Treated $\times$ 1940 $\times$ Married in $t$			1.520*** (0.4299)		-0.0430*** (0.0131)	
Dep. Var. 1930 Treated Mean	0.4819	10.39	10.39	0.1421	0.1421	
County fixed effects	Yes	Yes	Yes	Yes	Yes	
Observations	59,542	59,542	59,542	59,542	59,542	
Adjusted R <sup>2</sup>	0.05953	0.09277	0.46260	0.01673	0.04023	

*Notes:* All columns show results from estimating Equation (2) 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  $t$ . The outcome in Column (4) is an indicator for whether the woman is in the LF in year  $t$ . 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  $t$ . Standard errors are clustered at the county level. All specifications include county fixed effects.

*Put in numbers for results that you are currently just stating as directions (ex. 5.3)*

**Authors' reply:**

Thank you for this suggestion. We have updated the text throughout to include the numbers.

*Figure 1: Is the untreated mean already starting to move in 1930?*

**Authors' reply:**

Thank you for this question. We conducted a t-test of the means (of the county shares of white teachers who are married women) between the untreated and treated counties shown in Figure 1, and find no significant difference between them. However, we agree that in the figure, it appears that the means are beginning to shift right in both the treated and control states starting in 1930. This does not pose a problem in our difference-in-differences design, given the shift is similar in both treated and control states, in line with parallel pre-trends.

Interestingly however, the overall shift towards longer, more formal teacher training does help explain the potential increase in married women in teaching from 1920 to 1930. In describing the marriage bar prohibitions in Section 2, we note that the share of married women employed in teaching slowly rose across all treated and control states until 1930. We have added a footnote that gives a new possible interpretation of this shift, which we reproduce here:

One potential contributor to the slight increase in married women teachers across states by 1930 was the shift to longer, more formal accreditation. With most teacher training programs changing from local 2-year normal schools to state-approved 4-year teaching colleges by 1930, schools in need of qualified teachers may have chosen to retain/rehire trained and married women teachers as a result ([16]).

*Figure 2 is really compelling*

**Authors' reply:**

Thank you, we thought so too!

*Tables: Make number of decimal places consistent. Are the treated and control states significantly different for any of the variables in Table 1?*

**Authors' reply:**

Thank you for this suggestion. We have updated Table 1 to include p-values from testing for differences in means between our treated group and our preferred control group, the neighboring Southern states, which we reproduce below for your convenience.

Despite the overall economic and cultural similarities between our treated and control states, we find a few statistically significant differences in 1930, which we describe in the main text as follows:

There are a few small but notable differences between KY/NC and the neighboring states—in 1930, neighboring states were more urban and had higher LFP among married women, higher unemployment rates, and greater shares of single women in teaching—which warrant robustness exercises using controls and different empirical strategies.

These differences in 1930 are not inherently a threat to our difference-in-differences design. However, they do provide further motivation for using our border counties design, matched counties design, and for including specifications with additional controls.

**Table 1.** Summary of key county-level statistics by county group in 1930

	All (1)	South (2)	Treated (3)	Neighb. Sth. (4)	p-value ((3) &(4)) (5)
<b>Panel A: General County Statistics</b>					
Population (Thous.)	39.61 (2.427)	26.69 (1.322)	26.3 (2.061)	26.59 (1.736)	0.912
School-Age Pop. (Thous.)	10.15 (0.541)	7.92 (0.32)	7.998 (0.516)	8.012 (0.447)	0.984
Share Urban	0.214 (0.005)	0.155 (0.008)	0.134 (0.013)	0.173 (0.016)	0.059*
LFP of Married Women	0.102 (0.001)	0.136 (0.003)	0.092 (0.004)	0.118 (0.005)	0.000***
LFP of White Married Women	0.081 (0.001)	0.082 (0.002)	0.067 (0.003)	0.079 (0.003)	0.011**
Num. Children of Marr. Wom.	2.036 (0.007)	2.247 (0.012)	2.355 (0.029)	2.308 (0.02)	0.181
Unemployment Rate	0.054 (0.001)	0.042 (0.001)	0.042 (0.002)	0.047 (0.002)	0.062*
<b>Panel B: County Statistics on White Teachers</b>					
Students/Teachers	33.85 (0.26)	46.37 (0.466)	47.3 (0.934)	42.14 (0.723)	0.000***
Share Men	0.196 (0.002)	0.191 (0.003)	0.219 (0.006)	0.206 (0.006)	0.136
Share Single Women	0.636 (0.002)	0.615 (0.003)	0.604 (0.007)	0.62 (0.006)	0.075*
Share Married Women	0.168 (0.002)	0.194 (0.002)	0.177 (0.004)	0.174 (0.004)	0.561
<i>N</i> (Counties)	3100	944	220	320	

*Notes:* All statistics are measured using the full count 1930 census data, aggregated to the county level [22]. Panel A presents means and standard errors of county-level variables for the whole county population, including population in thousands, the percentage of the county population living in an urban area, the percentage of married women and white married women in the county between the ages of 18 and 64 who are in the labor force, and the average number of children for married women. Panel B presents means and standard errors of county-level variables related to teachers, including the white school-age population divided by the number of white teachers in a county, and the share of white teachers in a county that are men, unmarried women, and married women. Column (5) presents p-values from *t*-tests for differences in means between the Treated and Neighboring South counties.

*Figures A7 and A8 look weird with the SE, particularly given Table B1 Column (7)—could you put in sample sizes? Explain more about the dataset, the universe.*

**Authors' reply:**

Thank you for pointing this out. The original appendix figures you refer to showed the effects of the prohibitions on the likelihood that unmarried women teachers in  $t - 10$  (1) had a child in year  $t$ , and (2) were in the LF in year  $t$ .

We have made two changes. First, we have removed the results on the effects on overall LFP, because with our updated individual-level linked analysis, we actually find no evidence of unmarried women teachers being pushed out of the labor force all together. As a result, our results unambiguously suggest that the prohibitions increased women's overall LFP, even without running an additional regression.

Second, to clarify and clean up the presentation of the fertility results, we have removed the figure and replaced it with Supplementary Table B5 in which we examine not only the effects of the prohibitions on fertility, but also on occupational scores and mobility. We reproduce the table below.

Column (1) shows the fertility result. As you noted, our estimated effect of the prohibitions on the likelihood an unmarried woman teacher in  $t - 10$  has a child in  $t$  is small and noisily estimated, despite the sizable sample. However, it is clearer in the table that the magnitude of the coefficient is also relatively small, given that 48% of unmarried women teachers in 1920 have children by 1930. We therefore interpret this finding as a null result.

**Table B5.** Estimated effects of marriage bar prohibitions on fertility, occupational scores, and mobility

Dependent Variable:	Has child in $t$		Occupational Score in $t$		Moves state in $t$	
Model:	(1)	(2)	(3)	(4)	(5)	
Treated $\times$ 1940	-0.0208 (0.0135)	0.7275* (0.4080)	-0.0949 (0.2835)	0.0401*** (0.0093)	0.0593*** (0.0118)	
Treated $\times$ 1940 $\times$ Married in $t$			1.520*** (0.4299)		-0.0430*** (0.0131)	
Dep. Var. 1930 Treated Mean	0.4819	10.39	10.39	0.1421	0.1421	
County fixed effects	Yes	Yes	Yes	Yes	Yes	
Observations	59,542	59,542	59,542	59,542	59,542	
Adjusted R <sup>2</sup>	0.05953	0.09277	0.46260	0.01673	0.04023	

*Notes:* All columns show results from estimating Equation (2) 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  $t$ . The outcome in Column (4) is an indicator for whether the woman is in the LF in year  $t$ . 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  $t$ . Standard errors are clustered at the county level. All specifications include county fixed effects.

*Minutia: “prohibition of marriage bars” is difficult to parse the first time it comes up—is it the prohibition of marriage? Or are marriage bars being prohibited? (It’s the latter, but is there a way to make that immediately clear either with the use of a hyphen or some rephrasing?) My preference would be the longer, “laws that prohibited schools from discriminating against married teachers (prohibition of marriage bars)” or something similar the first time you introduce it*

**Authors’ reply:**

Thank you for this suggestion. This phrase is one that we went back and forth on a lot, since as you pointed out, it is difficult to express concisely! We have replaced the first sentence where we mention marriage bars in the introduction, which previously read:

This paper studies how married women’s labor supply in the early 1900s was affected by an important policy change: the prohibition of marriage bars in the teaching profession in two U.S. states in the 1930s.

with the following:

This paper studies how married women’s labor supply in the early 1900s was affected by an important policy change: the passage of laws that prohibited the use of “marriage bars” in the teaching profession in two U.S. states in the 1930s.

We hope that this adjustment makes the content clearer.

*p.4 What do you mean by “unconditional likelihood” that a married woman worked... unconditional on what?*

**Authors’ reply:**

Thank you for this question. As you raised in an earlier comment on Footnote 4, our usage of “conditional” and “unconditional” throughout the paper was sometimes confusing, so we have either removed the words (as we do in this case on p. 4) or altered the wording (as we did for Footnote 4) throughout the paper to clarify what we mean.

To answer your question: the original sentence on p. 4 reads:

The unconditional likelihood that a married woman worked as a teacher also increased by 17% 1930 and 1940 in treated counties relative to control ones.

What we meant in this case was that the share of married women who were teachers increased more in treated counties compared to control ones between 1930 and 1940. We used the word “unconditional” to indicate that the shares we compute do not “condition on” married women’s labor market status in 1930. However, in hindsight, this is an unnecessary distinction to make.

*Top of page 5, you note that the bulk of employment discrimination against married women*



*did in fact cease—I think you mean legislated employment discrimination or statutory employment discrimination? (Audit studies and lab experiments still show discrimination against married women, particularly those with children, for high level positions in the US and many positions in China.)*

**Authors’ reply:**

Thank you for pointing this out: you are correct that we meant to say statutory employment discrimination ceased. However, we have since removed the sentence (albeit for a different reason: the original sentence discussed a back-of-the-envelope calculation we had, which the editor suggested we remove).

*Top of page 7 is kind of repetitive and could be cut or consolidated with previous paragraphs (paragraph that discusses Goldin)—I think you’ve already said most of this by this point.*

**Authors’ reply:**

Thank you for this advice. We have cut down this introduction paragraph to spend less time discussing Goldin again, and add a sentence at the end highlighting our new findings on Black versus white women teachers. The new paragraph reads as follows:

This paper contributes to our understanding of the historical factors that led to the rise in women’s LFP in the U.S. throughout the 20th century. A significant body of work studies the effects of large-scale factors post-1950, including World War II [14, 2, 21], the introduction of oral contraceptives [13, 3], shifting cultural attitudes [7], and improved household technologies [15]. However, few papers study the factors that contributed to the initial rise pre-1950, which happened in spite of married women facing active, legal employer discrimination. One notable exception is Goldin (1988) [11], who documents the use of marriage bars in the early 1900s U.S. and explores firms’ economic justifications for using these discriminatory practices.<sup>15</sup> Motivated by Goldin (1988), we study how the prohibition of marriage bars impacted married women’s LFP in the first half of the century. Our findings suggest that the prohibition of marriage bars had sizeable effects on married women’s LFP in the short run, falling within the range of effect sizes of changes in the second half of the century like early access to the pill (which accounted for 8% of the rise in women’s LFP between the ages of 26-30, according to Bailey (2006) [3]) and improved household technologies (which accounted for over half of the later rise in women’s LFP, according to Greenwood et al. (2005) [15]). We also provide the first empirical evidence that marriage bars were not binding for Black women, as posited by Goldin [11].

---

<sup>15</sup>A few papers study marriage bars in other countries, including [20] and [19], which study the long-term effects of the marriage bar on teachers in Ireland.

*p.14 “counties in KY and NC had higher student-to-teacher ratios” compared to who?*

**Authors’ reply:**

Thank you for raising this clarifying question. The original sentence read:

Like the average Southern county, counties in KY and NC had higher student-to-teacher ratios, exhibited a low share of the population living in urban areas, and saw more children per married woman on average.

We meant to say that counties in KY and NC had higher student-to-teacher ratios compared to the average U.S. state. However, we have since removed this sentence and altered the paragraph following your previous recommendation to rearrange our description of the summary statistics in Table 1.

*What is “Sample 1” in Table 5?*

**Authors’ reply:**

Thank you for catching this typo! We have updated the table and the table notes to clarify that we use “Sample 1” to refer to the sample of women in our linked analysis who are unmarried and teaching in  $t - 10$ .

*Table B2 would be more clear I think if you used and (the upside down U) instead of |, since I think you’re saying that they’re a teacher in  $T$  and married in  $T$ , whereas when you have the conditional you’re thinking teacher in  $T+1$  given married in  $T$ . (What you have is correct, it’s just going against a convention and making it more confusing than it needs to be.)*

**Authors’ reply:**

Thank you for this suggestion. We have changed the labels in this table to avoid using the confusing “conditional” notation: instead, we now write that our outcomes of interest are whether a woman in  $t - 10$  is: Married in  $t$ , a Married Teacher in  $t$ , a Married Non-Teacher in LF in  $t$ , or Married and Not in the LF in  $t$ .

## References

- [1] Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson. Europe’s tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review*, 102(5):1832–1856, 2012.
- [2] Daron Acemoglu, David H. Autor, and David Lyle. Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury. *Journal of Political Economy*, 112(3), 06 2004.

- 
- [3] Martha J. Bailey. More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply. *The Quarterly Journal of Economics*, 121(1):289–320, 02 2006.
  - [4] Andriana Bellou and Emanuela Cardia. The great depression and the rise of female employment: A new hypothesis. *Explorations in Economic History*, 80:101383, 2021.
  - [5] Kasey Buckles, Adrian Haws, Joseph Price, and Haley Wilbert. Breakthroughs in historical record linking using genealogy data: The census tree project. Working Paper 31671, National Bureau of Economic Research, September 2023.
  - [6] Dora L Costa. From mill town to board room: The rise of women's paid labor. *Journal of Economic Perspectives*, 14(4):101–122, 2000.
  - [7] Raquel Fernández. Women, Work, and Culture. *Journal of the European Economic Association*, 5(2-3):305–332, 05 2007.
  - [8] T Aldrich Finegan and Robert A Margo. Work relief and the labor force participation of married women in 1940. *The Journal of Economic History*, 54(1):64–84, 1994.
  - [9] Price V Fishback, Michael R Haines, and Shawn Kantor. Births, deaths, and new deal relief during the great depression. *The review of economics and statistics*, 89(1):1–14, 2007.
  - [10] Emery M Foster, David T Blose, and Walter S Deffenbaugh. Biennial Survey of Education in the United States, 1930-1932. Bulletin, 1933, No. 2. Chapter I: Statistics of State School Systems for the Year 1931-32. *Office of Education, United States Department of the Interior*, 1933.
  - [11] Claudia Goldin. Marriage bars: Discrimination against married women workers, 1920's to 1950's. Working Paper 2747, "National Bureau of Economic Research", October 1988.
  - [12] Claudia Goldin. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton University Press, Princeton NJ, 2021.
  - [13] Claudia Goldin and Lawrence F. Katz. The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy*, 110(4):705–946, 08 2002.
  - [14] Claudia D. Goldin. The role of world war ii in the rise of women's employment. *The American Economic Review*, 81(4):741–756, 1991.
  - [15] Jeremy Greenwood, Ananth Seshadri, and Mehmet Yorukoglu. Engines of liberation. *The Review of Economic Studies*, 72(1):109–133, 2005.
  - [16] Charles Athiel Harper. A century of public teacher education: The story of the state teachers colleges as they evolved from the normal schools. 1939.
  - [17] Matthew J Hill. Love in the time of the depression: The effect of economic conditions on marriage in the great depression. *The Journal of Economic History*, 75(1):163–189, 2015.

- [18] Patrick Kline and Enrico Moretti. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. *The Quarterly journal of economics*, 129(1):275–331, 2014.
- [19] Irene Mosca, Vincent O’Sullivan, and Robert E Wright. The educational attainment of the children of stay-at-home mothers: evidence from the irish marriage bar. *Oxford Economic Papers*, 73(2):534–560, 2021.
- [20] Irene Mosca and Robert E Wright. The long-term consequences of the irish marriage bar. *The Economic and Social Review*, 51(1, Spring):1–34, 2020.
- [21] Evan K. Rose. The rise and fall of female labor force participation during world war ii in the united states. *The Journal of Economic History*, 78(3):673–711, 2018.
- [22] Steven Ruggles, 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.
- [23] Jessamyn Schaller, Price Fishback, and Kelli Marquardt. Local economic conditions and fertility from the great depression through the great recession. *AEA Papers and Proceedings*, 110:236–240, 2020.