Milestone 6.5

Molly Chiang

4/3/2020

Abstract

This is an extension of Marshall (2015), which shows the causal effect of additional years of schooling on voting conservative via regression discontinuty analysis on voting records of individuals before and after the British 1947 school-leaving age reform. This extension aims to see if the same effect is found in an analysis of a later education reform in England.

Introduction

This paper by John Marshall uses data from the results of the 1947 high school leaving age reform in Great Britain, to analyze how additional years in high school affect political preferences (Marshall 2015). This paper looked specifically in how additional years in high school effected voting for the Conservative Party (Marshall 2015). In 1947, Great Britain changed the high school leaving age from 14 to 15, this induced almost half the student population to stay in school for at least 1 or 2 more years (Marshall 2015). Data from the 10 British elections between 1947 and 2010 was then used to compare voters young enough to have been effected by the reform to those who were too old to have been effected, using regression discontinuity (Marshall 2015). Regression discontinuity is usually used for determining if a program/treatment is effective, and essentially is is a pretest-posttest program-comparison group design strategy (Trochim 2020). Regression discontinuity is unique in that individuals are assigned to one of two groups, just based on if they are on either side of a pre-determined cut-off (Trochim 2020).The results of data analysis revealed staying in high school for longer substantially increased likelihood to vote for the Conservative Party (staying one extra year increased probailitiy of voting Conservative by almost 12 percentage points) (Marshall 2015). This supports the previously studied fact that high school is extremely pertinent to political opinions later in life,

(could be two sentences and could have their own citations)

and that more education generally leads to higher income and thus voting more conservative (Marshall 2015). In addition, this significant finding indicates the education reform of 1947 may have had an even greater affect on politics and election results nationwide over many years than ever expected (Marshall 2015).

The first aspect of this project was a replication of Marshall's results. His code and data are publically available on the Harvard Dataverse. In order to replication Marshall's results, I ran his original code–with a few modifications–in stata. Marshall's figures were able to be replicated by running stata code, but replicating the tables was met with some difficulty. First, the rdrobust package has been updated since 2015 and thus some of the arguments Marshall used have since been deprecated (Sebastian Calonico 2017). I was able to replace old arguments with their 'updated' versions, according to an update from the package authors in 2017 (Sebastian Calonico 2017). Changing this argument allowedthe code to be run, but changed the values of the calculated coefficients by a bit. In addition, the code to go from raw output to polished table was not included in the replication code on the Dataverse, so I worked with Gov1006 Teaching Assistant Alice Xu to hard code the replication for table 1 (the main results table). All code for the replication is available in my github repo.[1].

After replication, I performed an extension on Marshall's original findings. (third and fourth paragraphs about what I did for extension and what I found)

Over the course of this paper, I will contextualize Marshall's 2015 paper in a literature review, dive more deeply into explaining my replication process, and show the results from my extension. The tables and

[1]Link to Github Repo

figures I replicated from Marshall's paper are also included at the end of the paper. (add more speculation at the end)

Literature Review

After the introduction, you will have a literature review, not dissimilar from the one in the paper you are replicating. (You do not get to assume that we have read the paper you are replicating. We haven't. So, if something is worth understanding about the literature, then you need to tell us, and in your own words.) You also need to closely review any relevant literature that has come out since the paper was published. (We will take off points if a simple Google scholar search brings up a relevant article which you should have mentioned.) Of course, if a lot of time has passed and/or this is a particularly active area of research, there may be dozens of relevant articles. You can't review them all. Pick the most important ones, especially those written by the same authors and/or using the same data and/or performing an analysis similar to your own extension.

Replication Process

I was able to replicate all of the figures in the paper in stata. The code I ran and the output figures are included in the appendix. In terms of the tables, I was able to run all of the code in stata almost directly from Marshall's replication code. Some aspects I had to change because some of the functions he used have been updated since he wrote the paper, namely, when using rdrobust, the bwselect option IK and the h() argument are no longer functional, and have been replaced with an updated version bwselect(mserd), which I used in all the code. This modification was necessary because of the changes that have been made to the rdrobust package since 2015, and did change my values a bit from Marshall's.

Although I was able to replicate the values of Marshall's tables by running his stata file (with modifications), I had a lot of trouble going from the raw results of the models Marshall ran to the polished tables I saw in his paper, as only code for the models were included in his .do file. Thus, I have included screenshots of the raw code and the relevant aspects of the output that would be included in the table in the paper. I hope to learn more basics of stata in order to compile and nicely display these raw outputs from Marshall's code.

For brevity, I did only included the screenshots from tables 1 and 2 and not table 3 (as it would be a lot of seperate screenshots to include), but I have included the table 3 code, which was run in the same way. Table 4 was not able to be replicated as there was no code for table 4 included in Marshall's replication code.

This paper relies heavily on rdrobust models, and the reference material for learning about this function and its package is from (Sebastian Calonico 2020).

My replication process was guided by advice from (Gary King 2000).

Extension

I believe there are two routes I could take with a proposed extension of the work in "Education and Voting Conservative: Evidence from a Major Schooling Reform in Great Britain" by John Marshall.

First, I could look at other factors that may have been influenced by the 1947 leaving age reform. This would involve doing a similar regression discontinuity analysis of other variables around the year 1947, but in R instead of stata. For example, did this reform lead to an uptick in college educated people? Or in another vein, how did the distribution of voting in all parties, liberal, labour, and conservative change after the reform, do we see a corresponding downtick in the other parties to go along with the observed uptick in conservative voting? In addition, I think that I could, using some of the adivce from RAOS, find more exciting/visually pleasing ways to display the results of these regression analyses.

On the other hand, I could use this data to look at the effect of other education reforms that have happened over the course of British history. There was a big education reform in England in 1988 ("History of

Education in England" 2020). This reform including introducing a national curriculum and corresponding national curriculum assesments, as well as modifying religious education ("History of Education in England" 2020). Although this reform did not result in increased years of school (so it would not be able to be compared to Marshall's main findings in that way), I think it could be interesting if this reform in content and uniformity of the British public education had any effect on voting preferences of the first classes to appreciate this reform (just as Marshall analyzed those who first appreciated the 1947 reform). Again, there could be issues in that this reform required more instiutional changes than an easy age cut-off, and thus likely took longer to implement, but the results could still be interesting. In addition, the results could shed light on Marshall's finidings and if they are specifically correlated with more years in school, or are perhaps a response to government intervention via school reforms.

In addition to the 1988 education reform, the school leaving age was increased to 16 years of age in officially in 1996 ("History of Education in England" 2020). This may pose an issue because Marshall's data only goes to 2010 (so we would have much more data on pre-reform individuals than post-reform indivudals), but would be very, very interesting addition to the findings of Marshall (2015), as it could really support the idea that additional years of school being associated with more conservative voting, if there was a similar uptick around this year. Additionally, if it did not show an uptick, it could show that perhaps Marshall discovered a unique effect of the 1947 British education reform, but not of more years of education in general.

Conclusion

5 paragraphs

Tables and Figures

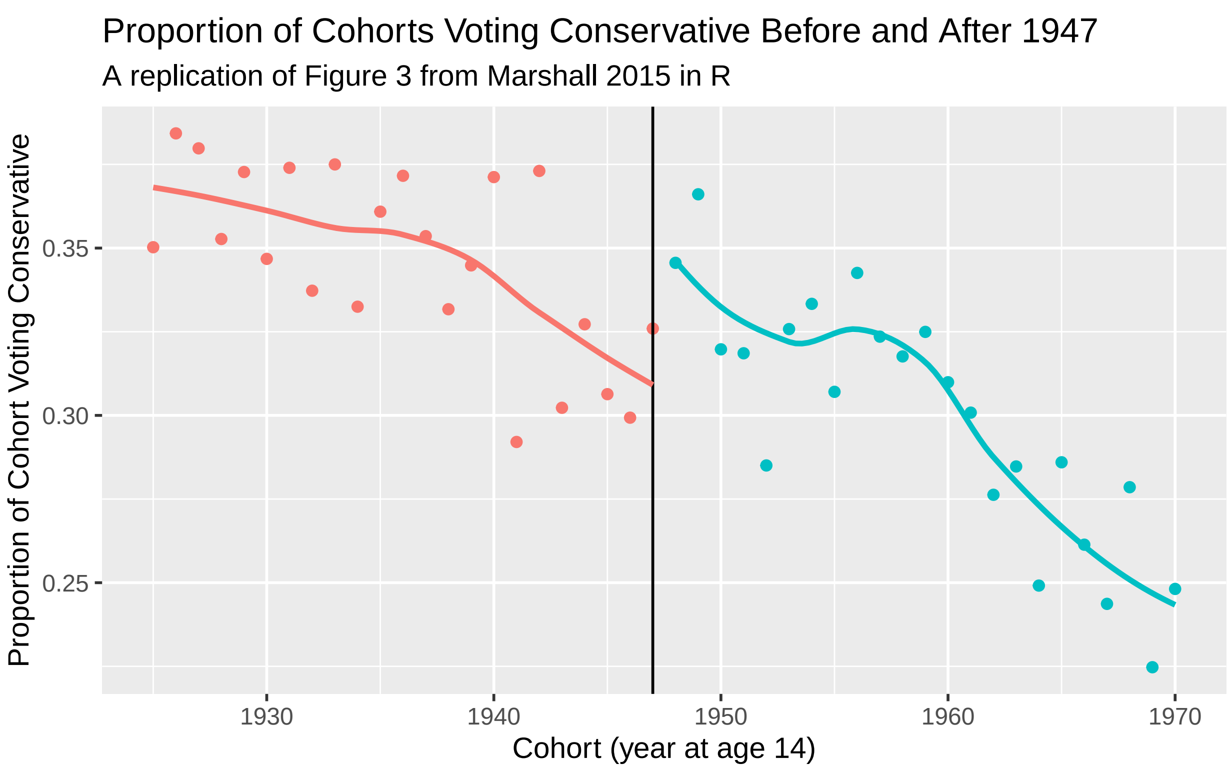

Figure 1: This figure is a replication of Figure 3 (Marshall 2015) in R instead of Stata, it illustrates the jump in proportion of a cohort, or class, of British individuals (classified by the year at which they were 14), after the education reform in 1947 which increased the year at which you could legally leave high school from 14 to 1

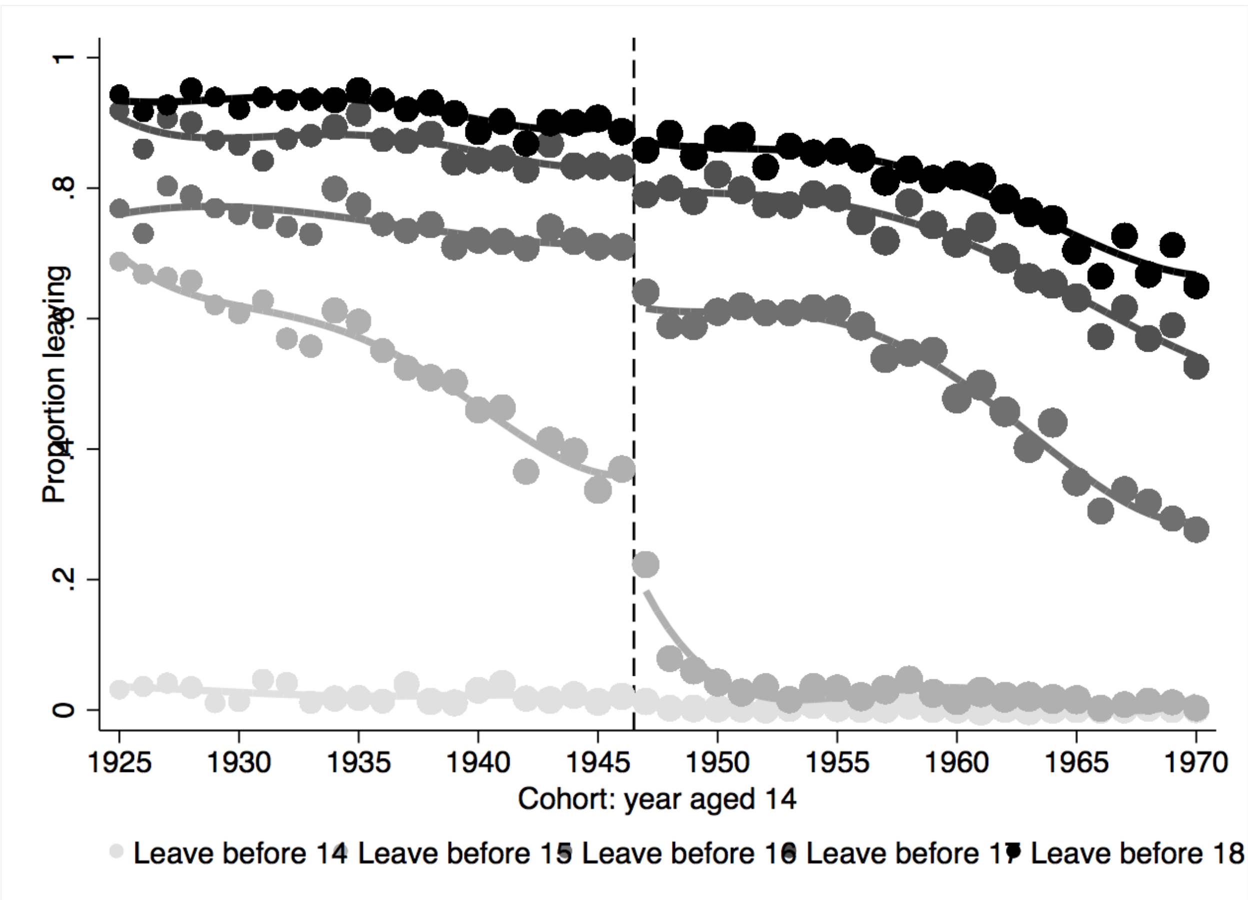

Figure 1: Marshall Figure 1

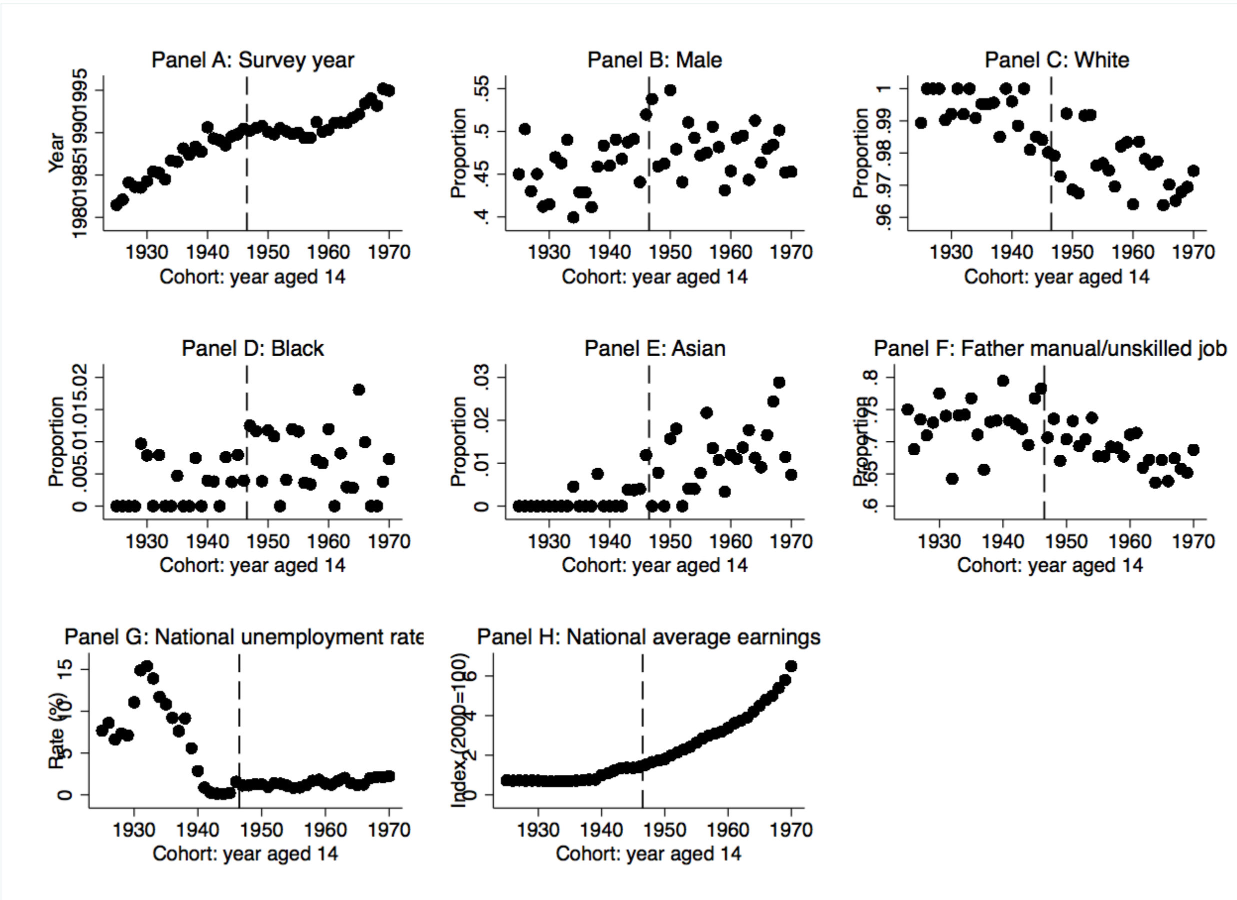

Figure 2: Marshall Figure 2

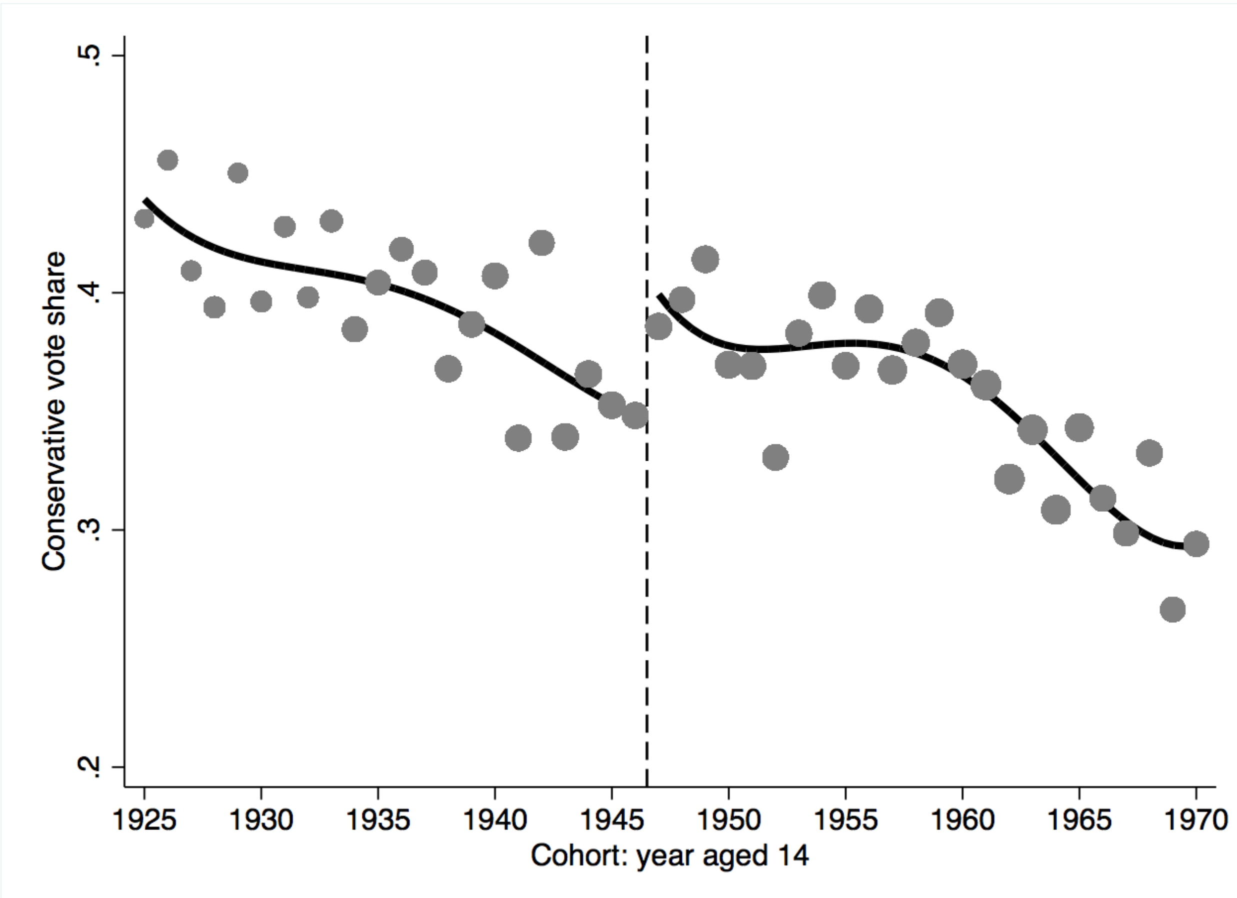

Figure 3: Marshall Figure 3

Table 1. Estimates of Schooling's Effect on Voting Conservative

	Years of Schooling LLR (1)	Attend University LLR (2)	Vote Conservative LLR (3)	Vote Conservative LLR IV (4)	Vote Conservative OLS (5)	Vote Conservative OLS (6)	Vote Labour LLR IV (7)	Vote Liberal LLR IV (8)
Post-1947 reform	.381***	.009	.044**					
	(.076)	(.013)	(.020)					
Years of schooling				.116**	.021***		−.071	−.021
				(.056)	(.002)		(.052)	(.043)
8th year of schooling						−.020		
						(.036)		
9th year of schooling						Baseline		
10th year of schooling						.126***		
						(.013)		
11th year of schooling						.213***		
						(.014)		
12th year of schooling						.289***		
						(.017)		
13th year of schooling						.306***		
						(.018)		
14th year of schooling						.281***		
						(.020)		
Outcome range	0–40	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1
Outcome mean	10.6	.11	.38	.38	.36	.36	.36	.19
Outcome standard deviation	1.86	.32	.49	.49	.48	.48	.48	.39
First stage *F*-statistic	25.4	.5		25.4			25.4	25.4
Observations	11,068	11,068	11,068	11,068	16,757	16,757	11,068	11,068

Figure 4: Original Marshall Table 1

Column 1

Coef.	Std. Err.
.4112	.09145

Variable	Obs	Mean	Std. Dev.	Min	Max
leave	11,068	10.58312	1.858667	0	40

Column 2

Coef.	Std. Err.
.03414	.01798

Variable	Obs	Mean	Std. Dev.	Min	Max
uni	11,068	.1145645	.31851	0	1

Column 3

Coef.	Std. Err.
.04548	.01829

Variable	Obs	Mean	Std. Dev.	Min	Max
leave	11,068	10.58312	1.858667	0	40

Column 4

Coef.	Std. Err.
.11581	.05539

Variable	Obs	Mean	Std. Dev.	Min	Max
leave	11,068	10.58312	1.858667	0	40

Column 5

Coef.	Robust Std. Err.
.021135	.0018537

Variable	Obs	Mean	Std. Dev.	Min	Max
con	16,757	.3587158	.4796379	0	1

Column 6

8	-.0200463	.0364474
10	.1257948	.0134744
11	.2126346	.0139522
12	.2891507	.0166584
13	.3061406	.0176145
14	.2810211	.0200969

Variable	Obs	Mean	Std. Dev.	Min	Max
con	16,757	.3587158	.4796379	0	1

Column 7

Coef.	Std. Err.
-.06892	.05082

Variable	Obs	Mean	Std. Dev.	Min	Max
lib	11,068	.1857608	.3889311	0	1

Column 8

Coef.	Std. Err.
-.00924	.04584

Variable	Obs	Mean	Std. Dev.	Min	Max
lib	11,068	.1857608	.3889311	0	1

Table 2. Schooling, Conservative Voting, and Income-Based Mechanisms

	Nonmanual Worker (below 60) (1)	Vote Conservative (below 60) (2)	Vote Conservative (60 or Above) (3)	Conservative Partisan (4)	Decided before Campaign (5)
A. Reduced form estimates:					
Post-1947 reform	.075**	.049**	.006	.039*	.041**
	(.030)	(.024)	(.029)	(.022)	(.019)
B. IV estimates:					
Years of schooling	.144***	.111**	.030	.092*	.103*
	(.055)	(.056)	(.153)	(.052)	(.053)
Outcome range	0 or 1	0 or 1	0 or 1	0 or 1	0 or 1
Outcome mean	.46	.38	.35	.37	.74
Outcome standard deviation	.50	.48	.48	.48	.44
Bandwidth	13.9	22.6	15.0	13.9	12.5
First stage *F*-statistic	28.1	28.3	2.7	26.4	23.4
Observations	6,086	10,152	4,589	9,711	9,510

Figure 5: Original Marshall Table 2

	Coef.	Std. Err.	Coef.	Std. Err.	Obs	Mean	Std. Dev.	Min	Max
Column 1	.10474	.0502	.15754	.06094	6,086	.4574433	.4982266	0	1
Column 2	.02718	.05559	.06984	.06284	10,152	.375591	.484299	0	1
Column 3	.02756	.04706	.81987	4.6518	4,589	.3497494	.4769427	0	1
Column 4	.03894	.0224	.08817	.05303	9,711	.369169	.4826046	0	1
Column 5	.04599	.0212	.1049	.05416	9,510	.7413249	.4379297	0	1

Figure 6: Replicated Table 2

Appendix (Code)

```
*** Figure 1: Trends in school leaving age
twoway (lpoly leave_l8 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs14) clwidth(thick) degree(4)) ///
  (lpoly leave_l8 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs14) clwidth(thick) degree(4)) ///
  (scatter meanleave_l8 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs14)) ///
  (lpoly leave_l9 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs11) clwidth(thick) degree(4)) ///
  (lpoly leave_l9 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs11) clwidth(thick) degree(4)) ///
  (scatter meanleave_l9 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs11)) ///
  (lpoly leave_l10 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs7) clwidth(thick) degree(4)) ///
  (lpoly leave_l10 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs7) clwidth(thick) degree(4)) ///
  (scatter meanleave_l10 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs7)) ///
  (lpoly leave_l11 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(gs5) clwidth(thick) degree(4)) ///
  (lpoly leave_l11 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(gs5) clwidth(thick) degree(4)) ///
  (scatter meanleave_l11 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gs5)) ///
  (lpoly leave_l12 yearat14 if yearat14<1947 & yearat14>=1925, lcolor(black) clwidth(thick) degree(4)) ///
  (lpoly leave_l12 yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(black) clwidth(thick) degree(4)) ///
  (scatter meanleave_l12 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(black)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion leaving) xtitle(Cohort: year aged 14) xline(1946.5,
lcolor(black) lpattern(dash)) xlab(1925(5)1970) ///
  legend(nobox region(fcolor(white) margin(zero) lcolor(white)) lab(3 "Leave before 14") lab(6 "Leave before 15") lab(9 "Leave
before 16") lab(12 "Leave before 17") lab(15 "Leave before 18") order(3 6 9 12 15) row(1))
```

```
*** Figure 2: Continuity graphs
capture by yearat14, sort : egen meanyear = mean(year)
twoway (scatter meanyear yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Year) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel A: Survey year, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g1.gph", replace

capture by yearat14, sort : egen meanmale = mean(male)
twoway (scatter meanmale yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel B: Male, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g2.gph", replace

capture by yearat14, sort : egen meanwhite = mean(white)
twoway (scatter meanwhite yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel C: White, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g3.gph", replace

capture by yearat14, sort : egen meanblack = mean(black)
twoway (scatter meanblack yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel D: Black, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g4.gph", replace

capture by yearat14, sort : egen meanasian = mean(asian)
twoway (scatter meanasian yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel E: Asian, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g5.gph", replace

capture by yearat14, sort : egen meanmanual = mean(fathermanual)
twoway (scatter meanmanual yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Proportion) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel F: Father manual/unskilled job, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g6.gph", replace

capture by yearat14, sort : egen meanurate = mean(urate)
twoway (scatter meanurate yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Rate (%)) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel G: National unemployment rate, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g7.gph", replace

twoway (scatter average_earnings yearat14 if yearat14>=1925 & yearat14<=1970, mcolor(black) msize(medsmall)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Index (2000=100)) xtitle(Cohort: year aged 14) xline(1946.5, lcolor(black) lpattern(dash)) ///
  legend(off) title(Panel H: National average earnings, color(black) size(medium)) xlab(1930[10]1970)
graph save Graph "g8.gph", replace

gr combine "g1" "g2" "g3" "g4" "g5" "g6" "g7" "g8", rows(3) cols(3) subtitle(, color(black) fcolor(white) lcolor(white)) graphregion(fcolor(white) lcolor(white) ifcolor(white)
ilcolor(white))
```

```
*** Figure 3: Reduced form
twoway (lpoly con yearat14 if yearat14>=1925 & yearat14<1947, lcolor(black) clwidth(thick) degree(4)) ///
  (lpoly con yearat14 if yearat14>=1947 & yearat14<=1970, lcolor(black) clwidth(thick) degree(4)) ///
  (scatter meancon14 yearat14 if yearat14>=1925 & yearat14<=1970 [weight=weight_14], msize(small) mcolor(gray)), ///
  graphregion(fcolor(white) lcolor(white)) ylab(,nogrid) ytitle(Conservative vote share) xtitle(Cohort: year aged 14) xline(1946.5,
lcolor(black) lpattern(dash)) ///
  yscale(range(.2 .5)) ylabel(.2[0.1]0.5) xlab(1925[5]1970) legend(off)
```

```
*** Table 1: Main estimates
rdrobust leave yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
sum leave if yearat14>=1933 & yearat14<=1961
rdrobust uni yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
sum uni if yearat14>=1933 & yearat14<=1961
rdrobust con yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust con yearat14, c(1947) fuzzy(leave) p(1) q(2) kernel(tri) bwselect(mserd)
sum leave if yearat14>=1933 & yearat14<=1961
areg con leave male white black asian sagesq-sagequart syearat14 syearat14sq syearat14cub syearat14quart, ro a(survey)
areg con ib9.leave male white black asian sagesq-sagequart syearat14 syearat14sq syearat14cub syearat14quart, ro a(survey)
summ con if e(sample)
rdrobust lab yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum lib if yearat14>=1933 & yearat14<=1961
rdrobust lib yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum lib if yearat14>=1933 & yearat14<=1961
```

```
*** Table 2: Raising social class, Heterogeneity by age (above 60), Become a Conservative partisan, and Decide before the electoral
campaign
rdrobust nonmanual yearat14 if age<60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust nonmanual yearat14 if age<60, fuzzy(leave) c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
sum nonmanual if age<60 & yearat14>=1934 & yearat14<=1960

rdrobust con yearat14 if age<60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust con yearat14 if age<60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum con if age<60 & yearat14>=1923 & yearat14<=1969

rdrobust con yearat14 if age>=60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust con yearat14 if age>=60, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum con if age>=60 & yearat14>=1932 & yearat14<=1962

rdrobust conpart yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust conpart yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum conpart if yearat14>=1934 & yearat14<=1960

rdrobust perm yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd)
rdrobust perm yearat14, c(1947) p(1) q(2) kernel(tri) bwselect(mserd) fuzzy(leave)
sum perm if yearat14>=1935 & yearat14<=1959
```

References

Reference material to create this bibliography comes from (J Allaire 2016)

Gary King, Jason Whittenberg, Michael Tomz. 2000. “Making the Most of Statistical Analyses: Improving Interpretation and Presentation.” Midwest Political Science Associatio. https://www-jstor-org.ezp-prod1.hul.harvard.edu/stable/pdf/2669316.pdf.

“History of Education in England.” 2020. https://en.wikipedia.org/wiki/History_of_education_in_England.

J Allaire, Jonathan McPherson, Yihui Xie. 2016. “Bibliographies and Citations.” https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html.

Marshall, John. 2015. “Education and Voting Conservative: Evidence from a Major Schooling Reform in Great Britain.” Southern Political Science Association. https://www-journals-uchicago-edu.ezp-prod1.hul.harvard.edu/doi/pdfplus/10.1086/683848.

Sebastian Calonico, Max H. Farrell, Matias D. Cattaneo. 2017. “Rdrobust: Software for Regression-Discontinuity Designs.” Midwest Political Science Associatio.

———. 2020. “Package ‘Rdrobust’.” https://cran.r-project.org/web/packages/rdrobust/rdrobust.pdf.

Trochim, Prof William M.K. 2020. “The Regression-Discontinuity Design.” https://socialresearchmethods.net/kb/regression-discontinuity-design/.