Gender, Education Reform, and Voting Conservative

Molly Chiang

4/3/2020

Contents

0.1 Abstract

Marshall (2015) shows the causal effect of additional years of schooling on voting conservative in his analysis og voting records before and after the British 1947 school-leaving age reform. Marshall's figures and the results of his tables were replicated in Stata, but an update in the rdrobust package lead to my modification of his bandwidth selection code, and thus slightly different coefficients. In an extension of Marshall's work I investigated how his observed effect differed between genders. Running rdrobust and creating regression discontinuity figures on just the male and female subsets of the data revealed the effect of more years of education increasing probabiliy of voting conservative was much stronger in women than men. This finding could complicate Marshall's argument that more education leads to higher income and then to more conservative political opinions and perhaps reveals something about the differing effect of education on men and women.

0.2 Introduction and Conclusion

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, and that more education generally leads to higher income and thus voting more conservative (Devereux and Hart

2010),(Meltzer and Richard 1981). 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 allowed the 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.

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

0.3 Literature Review

This paper is written as a response to mixed literature on education and voting preferences. On one hand, it has been suggested that education leads to more socially liberal attitudes. In 1959, Lipset proposed more education is assocaited with more liberal attitudes and support of democracy (Lipset 1959). Hyman and Wright support that analysis saying education leads to students thinking with a fundamentally liberal fashion (Hyman and Wright 1979).

On the other hand, Devereux and Hart showed in 2010–using similar data and the same regression discontinuity analysis design–that additional years of schooling as a result of the 1947 reform increases wage by 5-15% (Devereux and Hart 2010). In addition Meltzer and Richard (1981) have shown that higher wage earners prefer low income tax and government spending, policies aligned more closely with the British Conservative party (Meltzer and Richard 1981).

However, Marshall's paper was the first to directly address the causal effects of voting conservative and more years of education. This paper also used these past findings about the effect of education to help exaplin Marshall's findings.

0.4 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 was unable to go 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. The summary statistics for all of the tables (number

[1]Link to Github Repo

of observations, outcome mean, etc) were replicated perfectly. However, switching the bandwidth selection did result in changes to the results of the rdrobust functions. Some values differed more than others from Marshall's values after changing to bwselect(mserd), but it seems when fuzzy regression discontinuity was run, the values differed less.

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).

0.5 Extension

In my extension of "Education and Voting Conservative: Evidence from a Major Schooling Reform in Great Britain" by John Marshall, I investigated how his results differed when subsetting his data for males and females.

Creating regression discontinuity figures on just the male and female subsets of the data revealed the effect of more years of education increasing probabiliy of voting conservative was much stronger in women than men. In addition, then running rdrobust, the coefficient (which when running RDD is an indicator of causality) for the female subset was considerably higher than the coefficient of the male subset (with the both gender coeffecient being intermediate).

The RDD I ran for the extension was the same as used by Marshall. I set _____ to _____ - I set bwselect(mserd) for all - I used fuzzy(leave) for all - as fuzzy is used in instances where the assignment is not 100% yes or no - like in this case the government changed the law but it was not perfectly followed or enforced. - I selected the conventional coefficient and its corresponding se, as that is what Marshall used.

The RDD was run on a female sample of 15,661 individuals and a male sample of 13,745 individuals. Total (in the all-gender sample) there are 29396 individuals).

This finding could complicate Marshall's argument that more education leads to higher income and then to more conservative political opinions and perhaps reveals something about the differing effect of education on men and women.

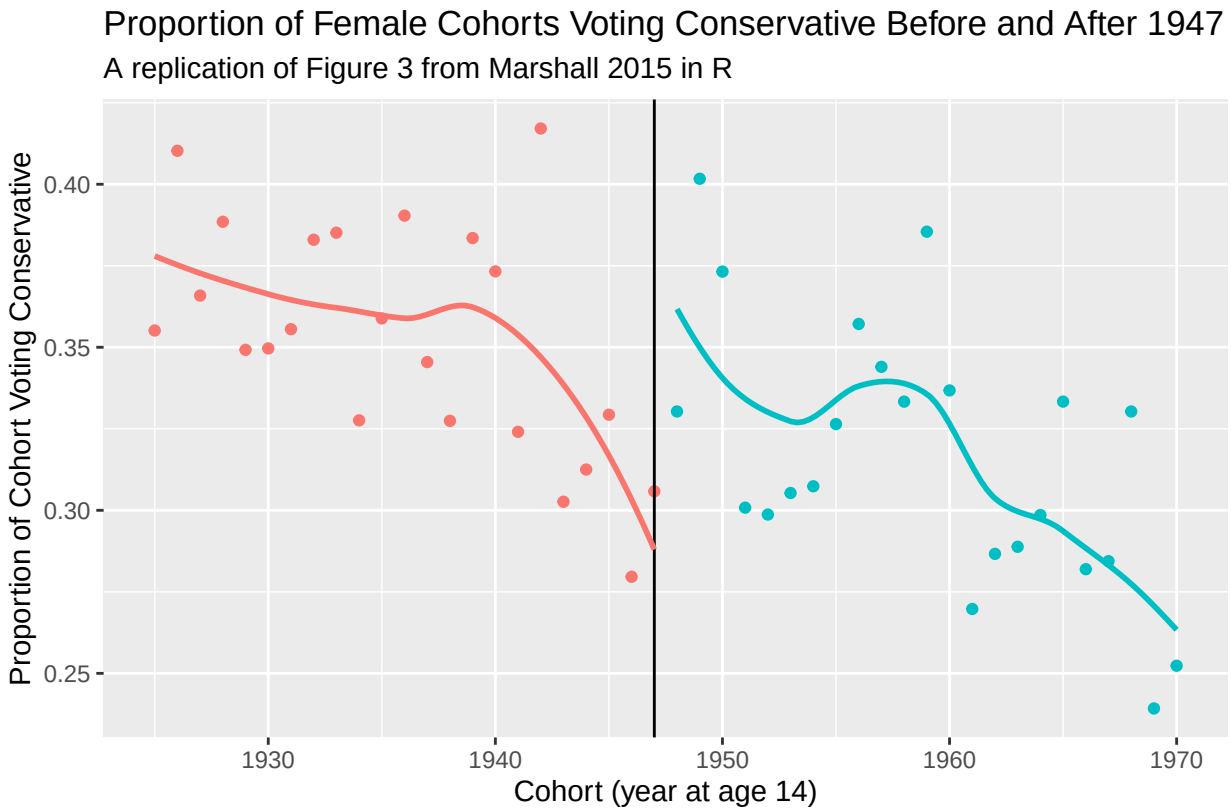

Figure 1:

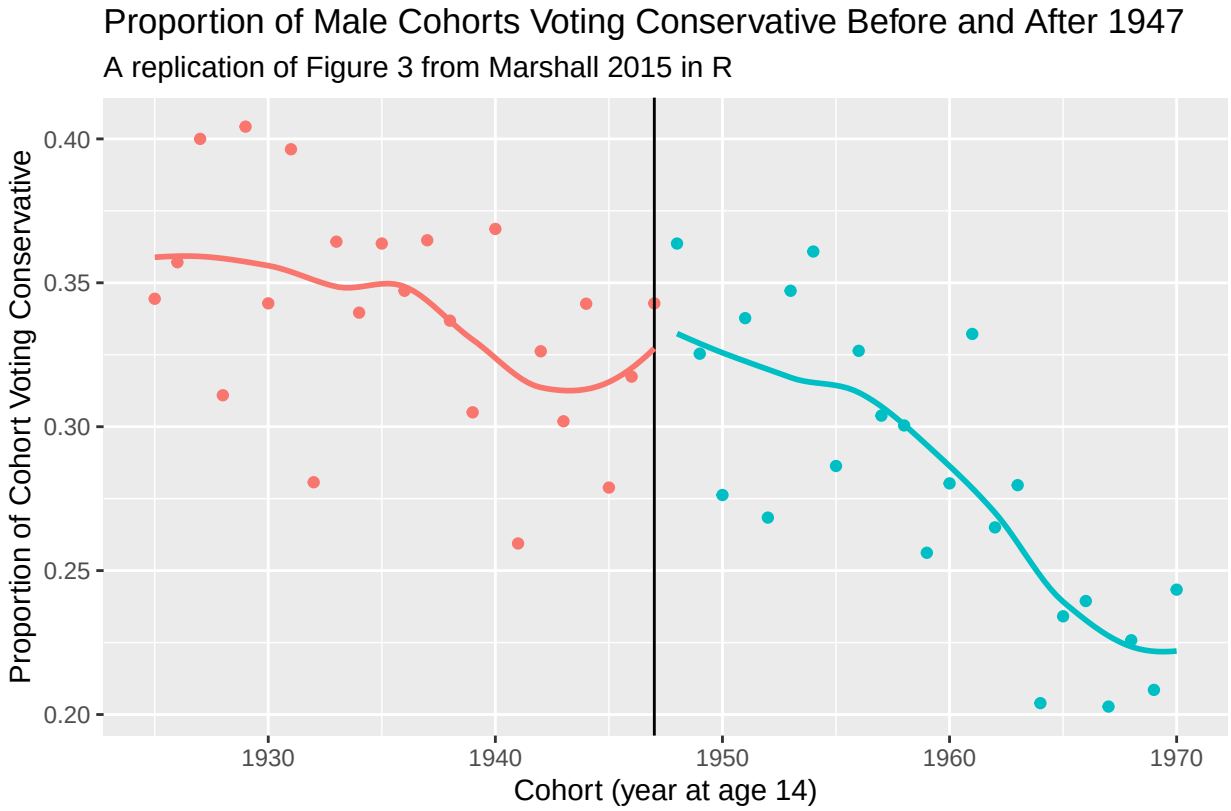

Figure 1:

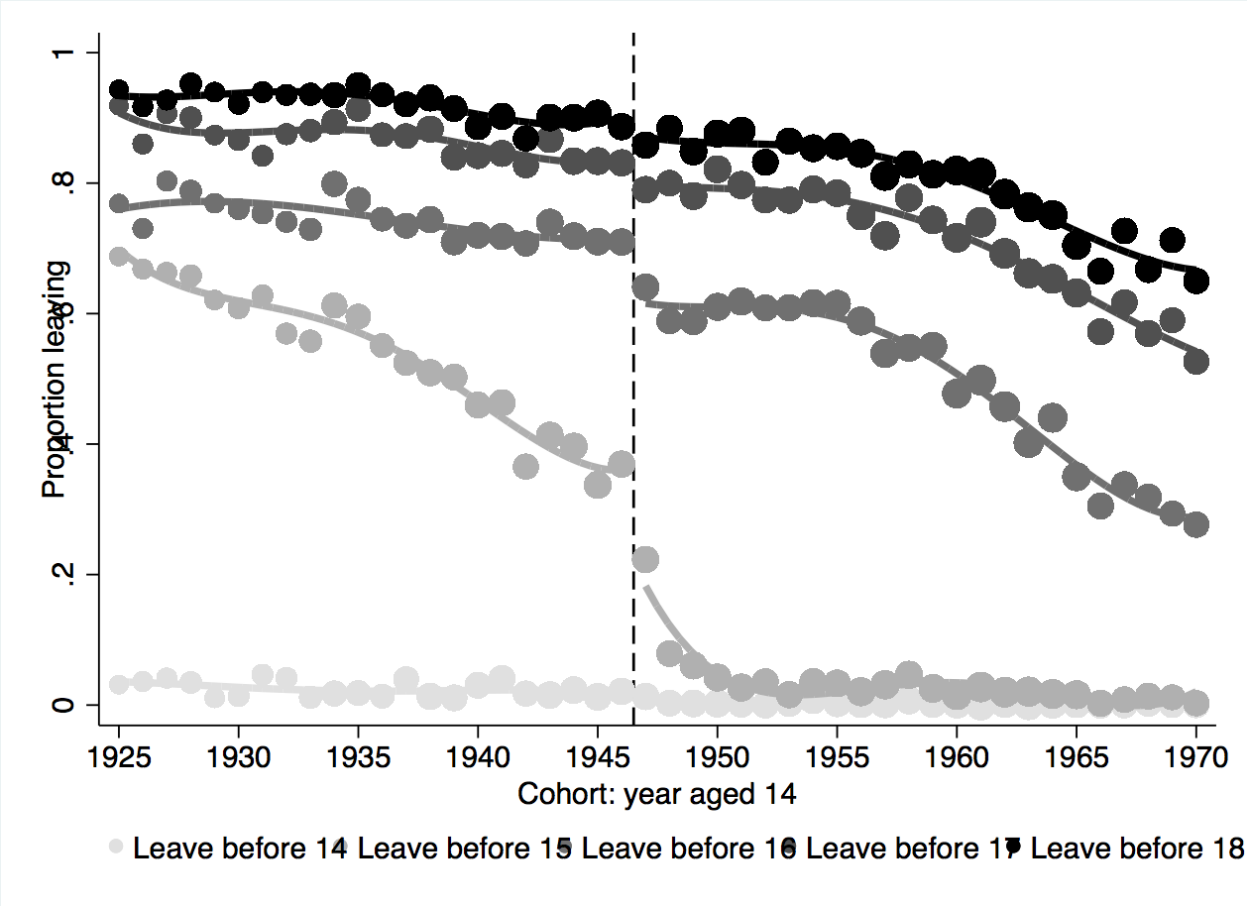

Figure 1: Marshall Figure 1 Replicated in Stata

Gender Breakdown of Schooling's Effect on Voting Conservative

Results of regression discontinuity analysis of a conservative vote around the 1947 school-leaving age reform in Great Britain

	All-Gender	Male	Female
Post 1974 Reform	Post 1974 Reform	0.045	0.039
0.053	Standard Error	Standard Error	0.019

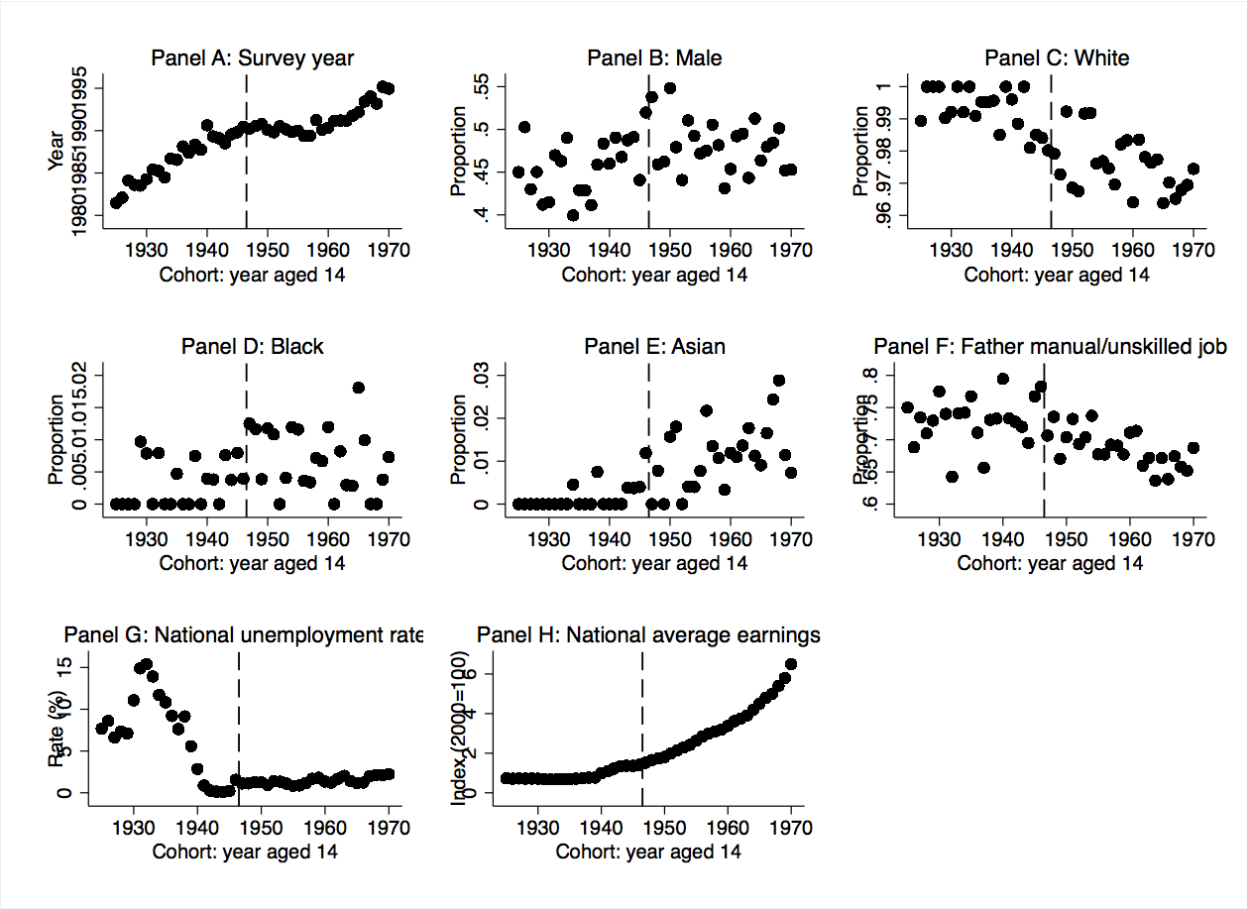

Figure 2: Marshall Figure 2 Replicated in Stata

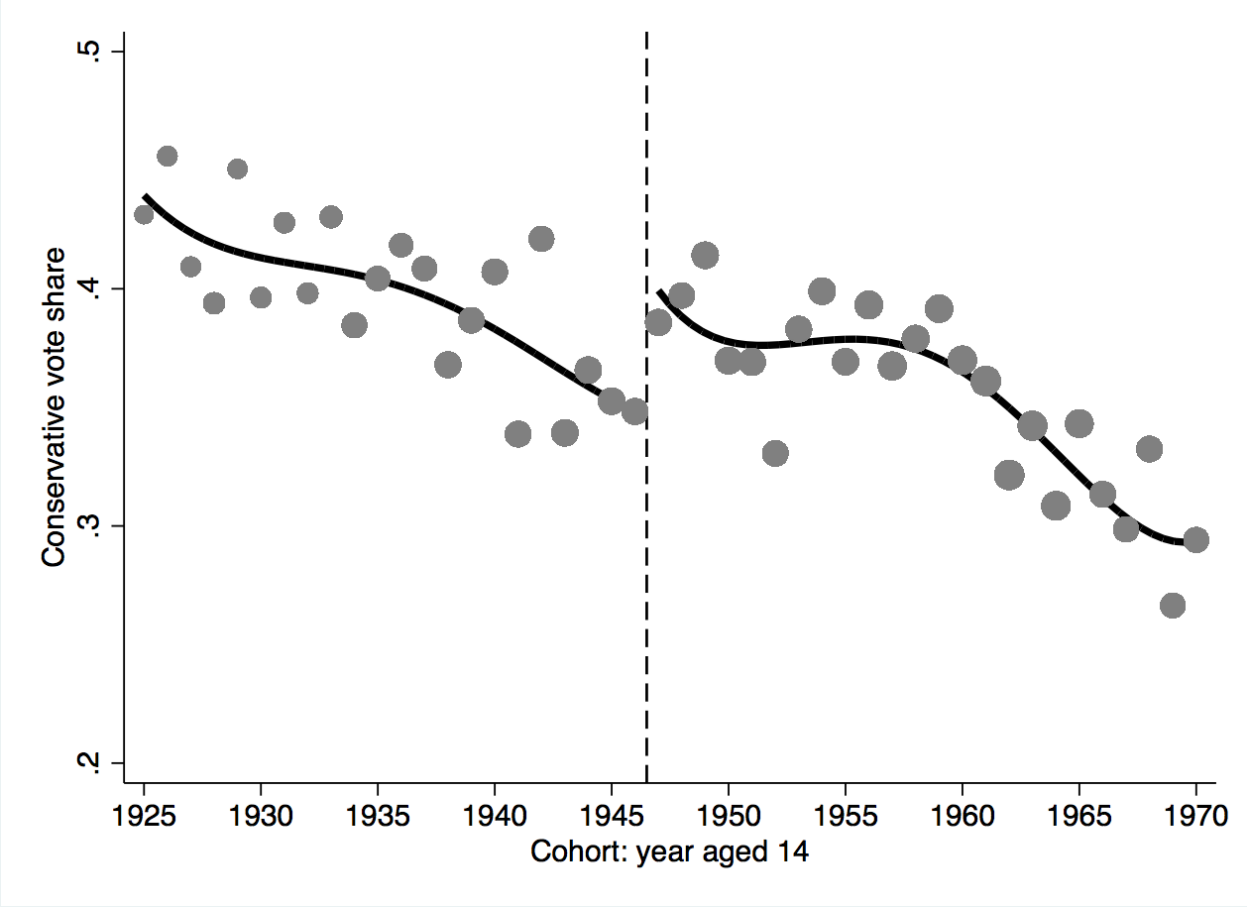

Figure 3: Marshall Figure 3 Replicated in Stata

0.6 Tables and Figures

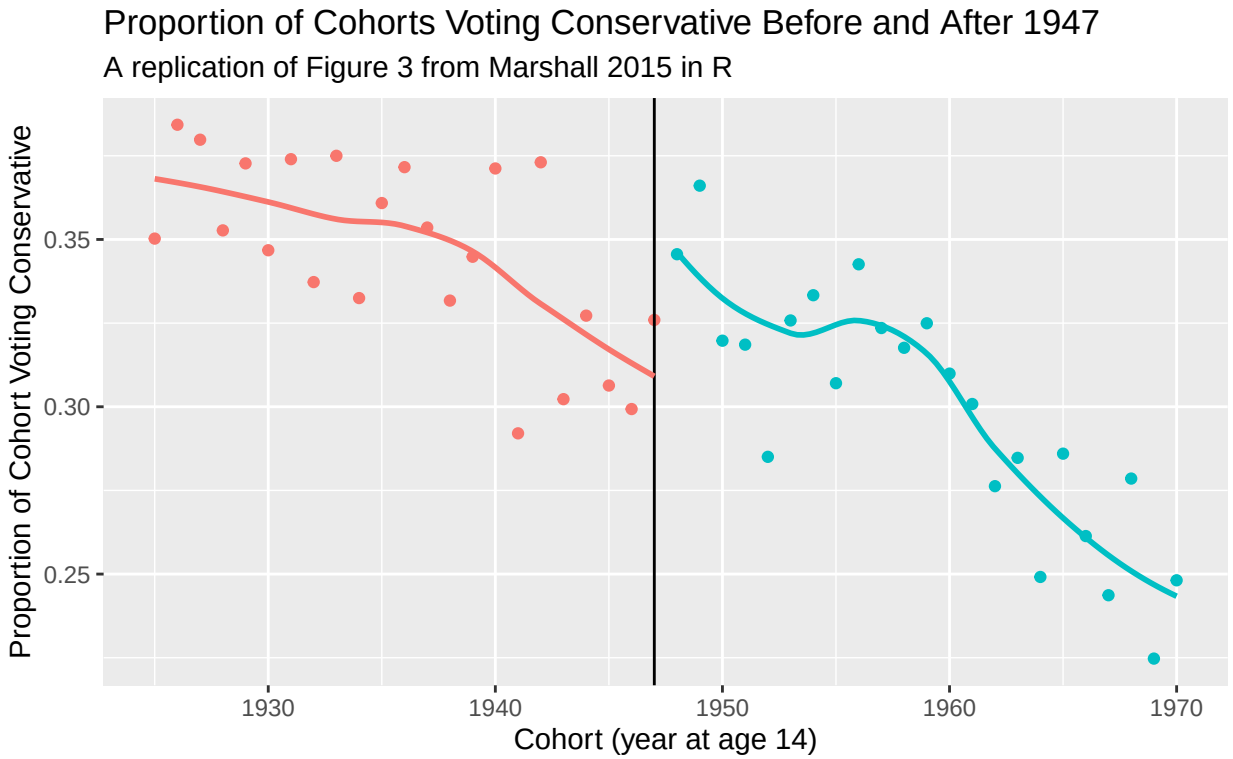

Figure 1: This figure is a replication of Figure 3 (Marshall 2015) in R instead of Stata, it illustrates the jump i proportion of a cohort, or class, of British individuals (classified by the year at which they were 14), after the education reform in 1947 which increased the year at which you could legally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

0.7 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
```

0.8 References

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

Devereux, Paul, and Robert Hart. 2010. "Forced to Be Rich? Returns to Compulsory Schooling in Britain." Economic Journal.

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.

Hyman, Herbert, and Charles Wright. 1979. "Education's Lasting Influence on Values." University of Chicago Press.

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

Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." American Political Science Review.

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.

Meltzer, Allan, and Scott Richard. 1981. "A Rational Theory of the Size of Government." Journal of Political Economy.

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

———. 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/.