Voting and Online Partisan Behaviour: a Regression Discontinuity Design

Lennard Metson

November 27, 2021

This extract has been adapted from a piece of assessed coursework from a second-year module called 'Research Design in Political Science'. The task was to plan and justify the design of an original research project on any topic in political science. Details of ethical and practical considerations, as well as details of robustness checks are contained in an Appendix which has not been attached.

Since the publication of Campbell et al.'s (1960) seminal work *The American Voter*, the causal mechanism through which partisan attachments are formed has been widely debated by political scientists. In particular, the correlation between age and strength of partisanship, established in the British electorate by Butler and Stokes (1969), has been explained by three competing mechanisms. These are summarised in *Figure 1.*

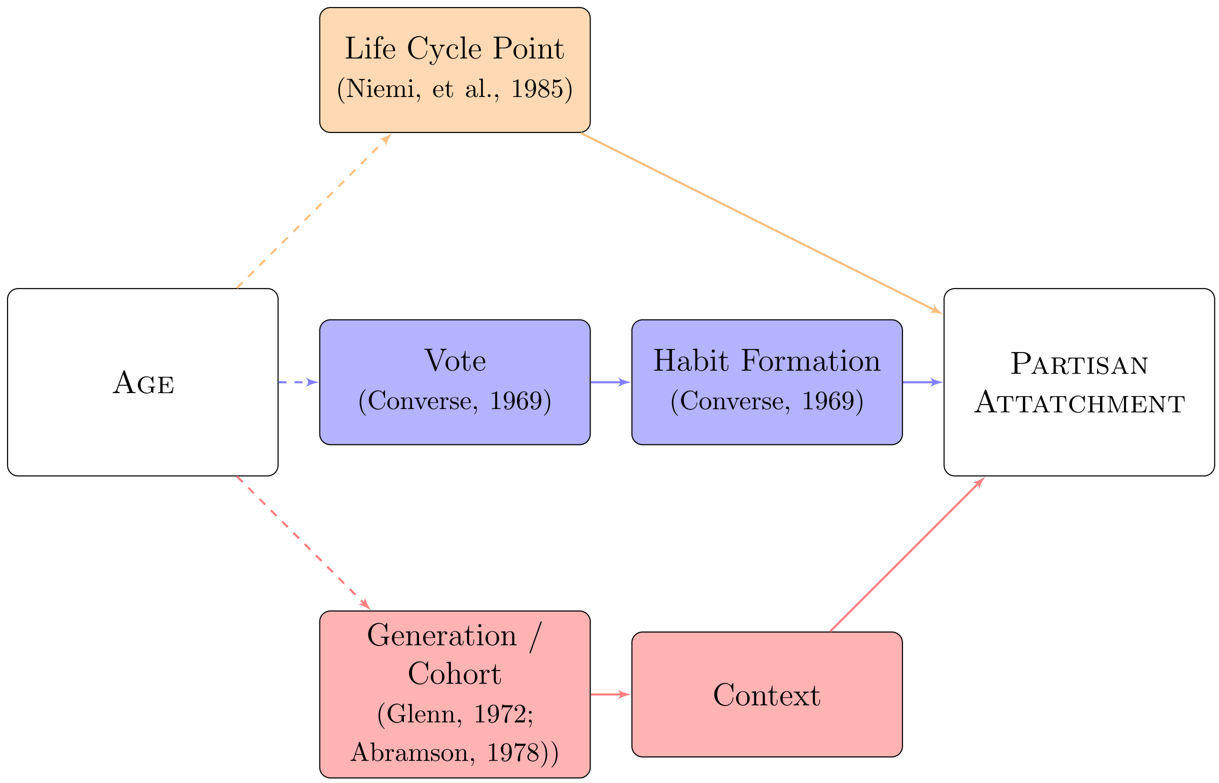

Figure 1: Possible mechanisms debated in the literature

According to Niemi, et al. (1985), the strength of somebody's partisanship is a result of their life stage. Following this argument, someone's experiences of policies and issues strengthen their partisanship as they get older. Therefore, older voters are more partisan than younger voters. Alternatively, as argued by Abramson (1978) and Glenn (1972), people of different generations have varying strengths of partisan attachment because they are socialised under different political contexts. Finally, Converse (1969) argues that voting is habit-forming and, thus, partisanship is strengthened through cumulative exposure to the act of voting. Therefore, each time somebody votes they should become more partisan. Tilley (2003) used a non-linear additive regression model to test the generational hypothesis and found that when controlling for length of electoral experience, generational effects are reduced to nearly zero (p. 341). Distinguishing the causal effects of the two remaining possible mechanisms is difficult as somebody's age determines both the length of their electoral experience, and their life stage. For this reason, the quasi-experimental 'Regression Discontinuity Design' (RDD) has proved useful in isolating the effect of a single act of voting (Dinas, 2012; Horiuchi et al., 2021). The presence of a cut-off exogenously divides people who are near the age of the eligibility cut-off into a 'treatment' group, who are eligible to vote, and a 'control' group, who are not. The design produces causal estimates because the division of a population who are of the same generation and life stage is as-good-as random. By comparing how frequently the subjects of each gorup express partisanship on social media, as a measure of partisanship, I seek to test the following hypothesis:

H_1: Citizens who can vote will express more partisan behaviour online.

The design I propose advances the causal examination of partisan formation in two ways. Firstly, it uses a measure of partisanship derived from subjects' online behaviour and secondly, it uses a variation of eligibility cut-offs to check the excludability assumption required in RD designs. Social media has become an increasingly important mediator of the relationship between voters and parties (Braby & Sältzer, 2021; McKelvey et al., 2014). To reflect the new ways in which voters interact with parties, I would use a textual analysis of subjects' Twitter feeds to measure the strength of their partisanship. Secondly, because several changes occur at the age of 18, and not just voting eligibility in General Elections changes at the age of 18, I would run a parallel analysis of the Scottish Parliamentary election where the eligibility age is 16. When these two elections run concurrently, we would be confident that the exludability assumption had been fufilled if we similar estimates at both cut-offs. My research design would thus provide novel causal evidence about the habit forming mechanism described by Converse (1969).

Sample and outcome measure

My population of interest is 13-22 year olds who would be eligible to vote in Scottish Parliamentary elections and General Elections at 16 and 18 respectively. Scotland has a variation of voting ages for different elections which means that the analysis can be carried out at two cut-offs simultaneously. It is possible that some of those in the General Election sample will have already voted at 16 in the Scottish elections. However, this would not create a biased result if they are balanced at either side of the 18-years-old cut-off. Following an approach used in Osmundsen, et al. (forthcoming), I would employ a survey firm to send out a survey to a representative sample of Scottish citizens to obtain their Twitter usernames, birthday and consent for their Tweets to be analysed. Demographic information would also be collected for balance checks. In order to ensure that my results are sufficiently well-powered, I would need a large sample size as, if partisanship formation from voting is a cumulative process, it is likely that the effect of each individual election will be small. With a sample of $n = 2000$, I would obtain results with 80% power and a p-value of < 0.05 for any effect size > 0.0075. Interpreted substantively, this is equivalent to the proportion of partisan Tweets in the treatment gorup being 0.79 percentage points higher than control.

My outcome variable is partisanship. Following Converse and Pierce (1985), I define partisanship as a stable pattern of positive behaviour towards a political party. This definition would include somebody tweeting or retweeting something positive about a particular party. Therefore, I would operationalise partisanship as posting or reposting positive content about a political party online. Ideally, I would use a semantic network method to determine the sentiment of Tweets that subjects make about different parties, to then determine their partisanship. However, doing so would likely be a breach of Twitter's API agreement which specifically bans inferring the political affiliation of individual users (Twitter, 2021). Therefore, I will ask the subjects which party they support and then count the number of Tweets they post or repost that reference said party. More partisan people should tweet more about their preferred party. Using an automated dictionary approach, I will define dictionaries for each party that contain stemmed words for referring to specific parties, as well as general terms such as "my party". Applying only the dictionary relevant for each subjects' preferred party, I will aggregate the proportion of a subjects' Tweets that contain a term from the dictionary. Converse and Pierce (1985) are clear that partisanship is not something that happens "overnight" (p. 145). However, if voting does cause partisanship, we should observe at least a small change in partisan behaviour. Of course, my measurement is subject to the assumption that people express their partisanship on Twitter. Recent research supports this claim. For Mosleh, et al. (2021) find that Twitter users are more than three times liklier to follow back accounts that match their partisan preferences. Finally, analysing samples of my computer-analysed data by hand will allow me to run a correlation agreement test to see how well this dictionary approach picks up partisanship.

Design assumptions, identification strategy and estimation

This study employs an RDD, applied to two separate eligibility cut-offs. The ability for people to vote (the treatment) is a function of somebody's age: 18 for General Elections and 16 for Scottish Parliamentary elections. Thus, eligibility creates a sharp cut-off which divides the population into those who can and cannot vote. Because people cannot select their exact birthdate, this division is as good as random assignment. This design comes with three core assumptions. Firstly, it assumes the sample is balanced around the cut-off because assignment is as-if-random. Of course, people cannot manipulate their age. However, to check that one group did not face significantly larger attrition than the other, I will carry out a McCrary density test between the two sides of the eligibility age. The second assumption is that the relationship between the assignment variable (age) and the outcome variable (partisanship) is linear around the cut-off. Forcing a linear regression model onto non-linear data can make it seem like there is a discontinuity in the outcome variable when there is not. According to data collected by Tilley (2003), the relationship is linear for my population of interest. To test linearity, I will check that my results are stable using different polynomial specifications of my basic regression equation. Finally, to fulfil the excludability assumption, no other relevant variables can change at the cut-off point. When the cut-off is 18, this is not the case. For example, at 18 people can leave education as well as take on debt or claim benefits. These are potentially relevant to the strength their partisanship. Therefore, I use the variation in voting age between elections to repeat my analysis at two separate cut-offs: 16 and 18. If other changes apart from voting do not have an effect, we would expect results of the same direction and significance at both cut-offs.

My identification strategy is based on voting eligibility laws. Eligibility to vote is the condition which assigns people to treatment or control (D_i). This assignment is determined by the cut-off (c_0) which is the age (X_i) at which someone becomes eligible to vote. Expressed formulaically: $D_i = 1$ if $X_i \geq c_0$ and $D_i = 0$ if $X_i < c_0$. Not everyone assigned to the treatment group will actually vote. These 'never-takers' are one of four hypothetical subgroups of compliance. It is very unlikely that people are never-takers for random reasons, and they are likely to exist in both assignment groups. Thus, I leave the never-takers in the treatment group, even if we can observe that they did not vote. Any other strategy, for example the local average treatment effect (LATE), would skew as-if random assignmen. Thus, I am estimating the intent-to-treat (ITT) effect of voting.

I will estimate the ITT of voting at each eligibility cut-off using linear regression. Assuming that the relationship between age and partisanship is linear within my population of interest, I will use a parametric strategy for my model specification. Where appropriate, this is preferred because it makes use of all the available data (Jacob et al., 2012). The basic OLS regression model I use is:

$$(1)\ Y_i = \alpha + \beta_1 * (X_i - c_0) + \beta_2 * D_i + \epsilon$$

where Y_i is the frequency of partisan posts, β_1 is the predicted linear effect of age on paritisanship, (X_i–c_0) is age normalised around the cut-off, and D_i is treatment status.

Thus, β_2 is the predicted effect which eligibility to vote has on partisanship. As my hypothesis is directional, I will use a one-tailed significance test. In order to reject the null hypothesis that eligibility to vote has no effect on partisanship, I would have to obtain a positive estimate of β_2 with a p-value < 0.05. I would run this regression analysis for both eligibilty cut-offs. Finding positive and significant results at 16 and 18 would provide evidence that voting causes stronger partisanship.

Conclusion

This paper has proposed an RD design that uses a computational measure of online partisan behaviour to test Converse's (1969) theory that partisanship strengthens cumulatively as citizens are exposed to the act of voting. The as-if-random nature of a voting age cut-offs makes a quasi-experimental causal estimation of the effect of voting itself possible. However, using RDD at the elgibility cut-off of 18 means that there are multiple changes to a citizen's rights at the same time, some of which might have an effect on their partisanship. For example, in the UK citizens also become eligible to take on economic rights and leave education at 18. If these changes also affect partisanship, the excludability assumption of quasi-experimental research is violated. To solve this problem, my research design takes advantage of elections that run concurrently in the same country with different eligibility ages. Scottish citizens are eligible to vote in Scottish Parliament Elections at 16 but must be 18 to vote in General elections. The only common change in rights at these two cut-offs is the right to vote. Thus, by comparing changes in partisanship at both ages at the same time, we can isolate the effect that voting has on partisanship. As the design proposed fulfils the assumptions of an RDD, finding a positive and significant increase in partisanship would provide causal support that the act of voting causes an increase in partisanship.

Bibliography

Abramson, P. R. (1978). Generational Replacement and Partisan Dealignment in Britain and the United States. *British Journal of Political Science, 8*, 505–509.

Braby, D., & Sältzer, M. (2021). *Server to the People - Dyadic Representation through Twitter in a Westminster System.* Unpublished Manuscript

Butler, D., & Stokes, D. E. (1969). *Political Change in Britain.* St. Martin's Press Inc.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). The American Voter (Unabridged). Wiley.

Converse, P. E. (1969). Of time and partisan stability. *Comparative Political Studies, 2*(2), 139–171. https://doi.org/10.1177/001041406900200201

Converse, P., & Pierce, R. (1985). Measuring Partisanship. *Political Methodology, 11*(3/4), 143–166. http://www.jstor.org/stable/41289338

Dinas, E. (2012). The Formation of Voting Habits. *Journal of Elections, Public Opinion and Parties, 22*(4), 431–456. https://doi.org/10.1080/17457289.2012.718280

Glenn, N. (1972). Sources of Shift to Political Independence: Some Evidence form a Cohort Analysis. *Social Science Quarterly, 53*, 494–519.

Horiuchi, Y., Katsumata, H., & Woodard, E. (2021). Young Citizens' Civic Engagement and Civic Attitudes: A Regression Discontinuity Analysis. *Political Behavior.* https://doi.org/10.1007/s11109-021-09698-7

Jacob, R. T., Zhu, P., Somers, M.-A., & Bloom, H. S. (2012). *A Practical Guide to Regression Discontinuity.* http://www.mdrc.org/publications/644/full.pdf

McKelvey, K., DiGrazia, J., & Rojas, F. (2014). Twitter publics: How online political communities signaled electoral outcomes in the 2010 US house election. *Information Communication and Society. 17*(4), 436–450.
https://doi.org/10.1080/1369118X.2014.892149

Mosleh, M., Martel, C., Eckles, D., & Rand, D. G. (2021). Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proceedings of the National Academy of Sciences, 118*(7), https://doi.org/10.1073/pnas.2022761118

Niemi, R., Powell, G. B. J., Stanley, H., & Evans, L. (1985). Testing the Converse Partisanship Model with New Electorates. *Comparative Political Studies, 18*, 300–322.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (forthcoming). Partisan polarization is the primary psychological motivation behind "fake news" sharing on Twitter. *American Political Science Review, Forthcoming.* https://doi.org/10.31234/osf.io/v45bk

Tilley, J. R. (2003). Party identification in Britain: Does length of time in the electorate affect strength of partisanship? *British Journal of Political Science, 33*(2), 332–344. https://doi.org/10.1017/s0007123403000152

Twitter. (2021). *Developer terms: More about restricted uses of the Twitter APIs.* https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases