

Arguing for a Negligible Effect

Seyoung Jung

I. Introduction

Social scientists pursue excavating relationships between different social factors. However, sometimes this leads to overemphasizing certain relationships. I think another important role of social science is to break some of the conventional beliefs or prejudices. There are quite a few articles theorizing that particular variables do not have an effect. According Rainey (2014), who reviewed research articles in the *American Political Science Review* and the *American Journal of Political Science* in 2011 and 2012, of the 75 articles that prevent explicit hypotheses, 22 (29%) use hypotheses of no effect (or a negligible effect) (Rainey 2014, 1083). Unfortunately, there is a relatively weak guidance to how to demonstrate evidence for a negligible effect. However, there is a growing attention to this topic, especially Wellek (2010) and Rainey (2014) provide a great book and an article addressing this issue.

Other than this topic being significant on its own, exploring it by reviewing suggestions of others is vital for my paper in progress. My paper is interested in the relationship between dual citizenship and the level of state attachment. I focus on measuring and describing the difference in means of two comparisons. My first analysis is comparing the levels of state attachment (attachment to the U.S.) of those who have one citizenship and those who have two. My second analysis is comparing the levels of state attachment to the two countries of dual citizens (e.g. to the U.S. and Ireland).

Relevant theories disagree on the impact of dual citizenship, which Staton et al. (2007) frames as a traditional versus transnational view. On one hand, traditionalists argue that citizenship should be a single, exclusive formal membership to one state and that dual citizenship would divide people's loyalty and attachments between two states (Geyer 1996; Huntington 2004; Renshon 2000; recited from Staton et al. 2007) On the other hand, transnationalists argue that multiple citizenship does not hinder political incorporation into the new country and that it is compatible with civic responsibility (Bloemraad 2004, Guarnizo et al. 2003). Between two competing theories, I expect to find a negligible effect, supporting the transnational view. From the online survey and implicit association test (IAT) I plan to conduct, I expect to find that the level of state attachment for the U.S. are similar between people with one citizenship and those with two citizenship. Also, I expect to find that dual citizens have similar levels of attachment for two states.

Although my current paper in progress relies on descriptive statistics, I plan to bring some designs that would allow me to make a causal inference in the near future. I think arguing for a negligible effect would be a critical aspect of such a claim. Merely suggesting a lack of statistical significance may not be sufficient for presenting "equivalence."

In order to study how to strengthen the argument of a negligible effect, this paper follows the guidelines suggested by Rainey (2014). Whereas he replicates Clark and Golder (2006), I replicate Krupnikov (2011) which Rainey (2014) references as well. First, I summarize, replicate, and analyze Krupnikov (2011). Second, I apply two suggestions by Rainey (2014) to improve the argument of Krupnikov (2011): 1) Two One-Sided Tests (TOST) using p-values and confidence intervals; and 2) perform many robustness checks.

II. Krupnikov (2011)

Summary of the paper

The study examines the effect of negative campaigning on voter turnout. Previously, there have been competing theories with conflicting empirical results: demobilization, mobilization, and no effect. The findings of the paper support that overall negativity has a null effect, with the exception of late targeted negativity. She establishes three models examining overall negativity, negativity by target (negativity about disliked candidate and liked candidate), and negativity by both target and timing (after the individual selected a candidate). Her three hypotheses are: 1) overall negativity has no effect on turnout; 2) negativity by target has no effect on turnout; and 3) negativity by both target and timing has demobilizing effect.

She uses seven presidential elections from ANES 1976 to 2000, merged with ad data from John Geer's database of presidential campaign advertisements (Geer, John, Presidential TV Ads 1960-2000). In addition to measures of negativity, she includes 23 controls including education and partisanship. She uses logistic regression with clustered standard errors. She first tests three hypotheses, then also presents change in likelihood of turnout due to an increase in negativity.

Replication of the findings

I replicate her findings of testing hypothesis 3, which shows that negativity by both target and timing has demobilizing effect. The replication of the findings are equivalent to the Krupnikov (2011) Table 4 Model 3.

| | |
|---------------------|-----------------------|
| Dependent variable: | turnout |
| negaboutdislike | -1.108 (0.771) |
| negaboutlike | -1.352* (0.788) |
| income | 0.831*** (0.200) |
| education | 1.719*** (0.191) |
| age | 0.025*** (0.003) |
| unemployed | -0.119 (0.185) |
| PIDStrength | 0.466*** (0.145) |
| AffectPID | 0.444* (0.248) |
| care | 0.125 (0.102) |
| AffectPRES | 0.554*** (0.205) |
| lnYears | 0.083** (0.038) |
| Church | 0.867*** (0.115) |
| homeowners | 0.477*** (0.100) |
| working | 0.157 (0.113) |
| contacted | 0.753*** (0.121) |
| external | 0.260** (0.112) |
| internal | 0.316*** (0.115) |
| interest | 0.321*** (0.070) |
| media_index | 0.390*** (0.102) |
| married | 0.205** (0.098) |
| black | 0.149 (0.135) |
| southern | -0.460*** (0.094) |
| hispanic | -0.006 (0.180) |
| gender | 0.085 (0.092) |
| closeness | 0.187* (0.099) |
| governors | -0.296** (0.122) |
| primaries | -0.085 (0.102) |
| volume2 | -0.334 (0.381) |
| dummy1988 | 0.079 (0.159) |
| dummy2000 | 0.324 (0.243) |
| dummy1992 | 0.822** (0.402) |
| dummy1996 | 0.308 (0.244) |
| Constant | -2.925*** (0.906) |
| Observations | 6,214 |
| Log Likelihood | -1,854.233 |
| Akaike Inf. Crit. | 3,774.467 |
| Note: | p<0.1; p<0.05; p<0.01 |

I replicate her findings of change in likelihood of turnout due to an increase in negativity. The replication of the findings are equivalent to the Krupnikov (2011) Table 3 Elections 1976 to 2000.

| est |
|------------------|
| 0% to 20% -0.011 |
| 0% to 40% -0.025 |
| 0% to 60% -0.041 |

Analysis of the paper

The strongpoint about this article is that it attempts to specify the scope condition in which negativity plays a role. I think finding such a condition made her argument more sophisticated rather than depending solely on overall null effect. Along the same line, I think using three hypotheses and their models was a good method demonstrating the effect in each condition. For example, I think comparing the negativity effects on liked candidates and disliked candidates were categorically good comparisons, but it is still ambiguous when individuals actually made selection. Krupnikov admits that and states that she generalized the selection time for the sake of translating the main theoretical prediction.

On the other hand, this article also has a few weaknesses. First, the fact that this is an observational study involves some caveats. One is the causal effect, meaning that we cannot be sure whether the negativity actually caused changes in turnout. For instance, there is no baseline information on what would have happened if individuals had not received negativity. Even though negative campaign advertisements were externally imposed, there may be other confounders that affect both negativity and turnout. This leads to another weakness which is including many controls. For the sake of causal inference, experimental design would be ideal in terms of taking care of the problem of observed and unobserved confounders. This is already addressed in the

article that there are existing researches using experimental research in psychology (Dholakia and Bagozzi 2003; Orbell and Sheeran 2000; Simon et al. 2001). Her rationale for observational research was to show that a nationally representative data can also show the result that is consistent with experimental research. However, I am still critical to using too many controls. On this issue, Achens warns the dangers to using too many controls, which creates too much noise and lacks substantive theoretical foundations and justification of adding controls. In the case of Krupnikov (2011), the only justification the article gives is that they are considered as traditional determinants of turnout (Krupnikov 2011, 802). Achens presents A Rule of Three (ART) which states that a statistical specification with more than three explanatory variables is meaningless (Achen 2002, 446).

Second, Krupnikov (2011) argues that there is no effect of late negativity to disliked candidates. Even though I agree with the thought of seeking for a null effect, it should present supporting evidences, which I will address once again in the last section.

The following are the analyses of the operating characteristics of the procedures, including p-values and confidence intervals using randomized based test, bias and consistency of estimators, and type I error and power of the test. This article uses logistic regression and I am curious about her choice. I do not see using logistic regression here as problematic. Freedman (2008) argues that randomization does not justify logistic regression and warns that it may be biased. I checked bias and power of the test, and found that it is unbiased and highly powered. However, rather than automatically resorting to logistic regression because of the use of dichotomous outcome variable, there needs to be a justification of why she sees voter turnout as a variable that would follow Bernoulli distribution. Even though at first glance, voter turnout following Bernoulli distribution seems plausible, but theoretical and empirical evidence would strengthen her argument.

P-values and confidence intervals

I calculate p-values based on Fisher's approach. I create a null world and see how unusual it would be to observe the coefficient. I also checked whether the distribution of no effects follow normal distribution, comparing it with a normal distribution curve and using qqplot. As a result, the distribution follows normal distribution, which means that central limit theorem is working. The p-value based on Fisher's approach was 0.3397 which was a bit larger than the one article used. I also calculated confidence intervals using the randomized based test, which gave me a range of -1.394299 to 1.269258.

Bias of estimators

To check bias, I created my truth as -0.2987017, using the coefficient from the glm function used in the article. Using the truth, I created a world of known effect, which has a distribution of possible coefficients. I compared the average coefficient with tau, to assess whether my estimator is biased or not. As a result, -0.2985196 was very similar to tau.

Consistency of estimators

To check consistency, I changed in number of observations. First, I created a smaller data sample, decreasing it from 6214 to 4000. Then, I created a larger data sample, enlarging it from 6214 to 10000. As a result, from 4000 to 6214, it became further from the truth, which is not consistent. On the other hand, from 6214 to 10000, it did become closer to truth, which is consistent.

Type I error of tests

To get Type I error, which is incorrect rejection of a true null hypotheses, I create a null world by breaking the relationship between independent variable and dependent variable. I conduct 10000 simulations, which means that out of 10000 cases, we want p-values smaller than 0.05 to be about 500 times. In this case, there are 494. In the world of no effects, p-values smaller than 0.05 are those cases that reject the null hypothesis even when it might be true. We would want our type 1 error to be smaller than 0.05. In this case, our type 1 error, 0.0494 is smaller than 0.05.

Standard errors

Standard error tells us the extent to which estimates will vary across all possible random assignments. I estimated the standard error of the p-value, taking into account type 1 error. In this case, the standard errors are very small close to 0.

Power of tests

I calculated power of test, which tells us how much the test avoids Type II error, which is incorrect retaining of a false null hypothesis. I used the truth and created a world of known effect, and conducted 10000 simulations, which means that out of 10000 cases, we want p-values smaller than 0.05 to be about 500 times. In this case, there are 517. The power of this test is 94.83%, which is very high.

III. Strengthening the argument

Propose changes in the approach to statistical (and/or causal) inference

As mentioned above, Krupnikov (2011) argues that there is no effect of late negativity to disliked candidates, but does not define the criteria of supporting evidences. On this issue, Wellek (2010) suggests specifying beforehand the size of the parameter interpreted as the substantively meaningful effect. Theoretically, Krupnikov (2011) could have included how big of a coefficient would be considered as substantively meaningful. Statistically, Rainey (2014) suggests Two One-Sided Tests (TOST) using p-values and confidence intervals, and performing many robustness checks. I take his advice and attempt them as the following.

Two One-Sided Tests (TOST): P-values

I made two hypotheses, one is alternative hypothesis and the other is null hypothesis. My alternative hypothesis is that the treatment effect is between $-.3$ and $.3$. (arbitrarily chosen size of the parameter interpreted as the substantively meaningful effect). My null hypothesis is that the treatment effect is either $-.3$ to negative infinity or $.3$ to infinity. I calculate two p-values using the alpha as $.1$ instead of conventional $.05$, making it more challenging for arguing negligible effect. I use the bigger one between the two one-sided p-values, also making it more challenging. In this case the p-value is $1.647578e-191$, it is small enough to reject the null, which supports a negligible effect.

Two One-Sided Tests (TOST): Confidence Intervals

As Krupnikov (2011) examined the change in likelihood of turnout due to an increase in negativity, I add confidence interval to the results. I use two one-sided tests, which makes it more challenging to argue for a negligible effect.

| | est | lwr | upr |
|-----------|--------|--------|-------|
| 0% to 20% | -0.011 | -0.018 | 0.002 |
| 0% to 40% | -0.025 | -0.045 | 0.005 |
| 0% to 60% | -0.042 | -0.085 | 0.007 |

Perform many robustness checks

Rainey (2014) replicated Clark and Golder(2006) and suggested performing many robustness checks (Rainey 2014, 1089). He conducted 6 robust checks. I referenced his replication code and tried them with Krupnikov (2011). Since the data of Clark and Golder (2006) was cross-sectional data across different countries, I omitted two cross-sectional tests. I conducted four robust checks, including standard error estimate, linear regression, robust regression, and bootstrapping method.

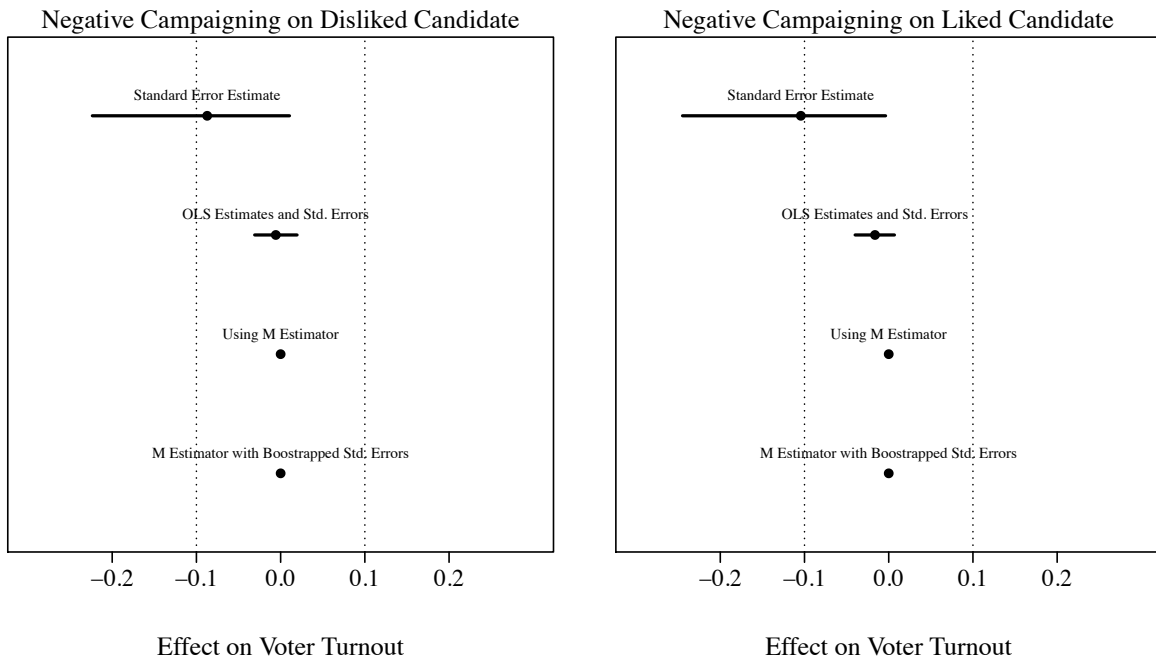
Late Negativity on Disliked Candidate

| | 5% | 50% | 95% |
|---|----------|----------|---------|
| Standard Error Estimate | -0.224 | -0.087 | 0.011 |
| OLS Estimates and Std. Errors | -0.031 | -0.006 | 0.020 |
| Using M Estimator | -0.00000 | -0.00000 | 0.00000 |
| M Estimator with Bootstrapped Std. Errors | -0.00000 | -0.00000 | 0.0000 |

Late Negativity on Liked Candidate

| | 5% | 50% | 95% |
|--|----------|----------|----------|
| Standard Error Estimate | -0.245 | -0.104 | -0.004 |
| OLS Estimates and Std. Errors | -0.040 | -0.016 | 0.007 |
| Using M Estimator | -0.00000 | -0.00000 | 0.00000 |
| M Estimator with Boostrapped Std. Errors | -0.00000 | -0.00000 | -0.00000 |

I plot the findings of robust checks in the form of confidence intervals. The plot shows that all of the confidence intervals are within the range of -.3 and .3. (arbitrarily chosen size of the parameter interpreted as the substantively meaningful effect). This supports the argument of negligible effect of negative campaigning.



IV. Conclusion

Replicating and analyzing Krupnikov (2011) was meaningful in numerous ways. Substantively, this paper provides great insight on how to observe individuals' perception and behavior. The author brings in studies of psychology and pays attention to how individual's decision process consists of selection phase and post-selection phase. She focuses on the latter, in which an individual has already decided on liked and disliked candidates. She observes individuals' response to external stimuli, which in this case was negative campaigning, and finds that responses are different according to timing and target. Personally, as a student of political psychology, I found psychology studies very interesting and informative. Nevertheless, through reviewing this article, I found that adopting psychology theories are not only useful for substantive reasons, but also for specifying hypotheses and models.

From the methodological perspective, reviewing this paper provides opportunity of thinking about observational studies and how to present a null effect. As observational studies have purposes and goals of their own, it would be a poor advice to recommend changing the study to experimental design. However, conducting observational studies would imply that there are many decisions and justifications to make. Although Krupnikov (2011) does a great job in explaining decisions she makes and acknowledges caveats, I think there can be at least one more change or justification made, decreasing the number of controls.

Krupnikov (2011) also does a great job in positioning a null effect. Since there have been debates on the effects of negativity for a while, she does not just present another finding of either mobilizing effect, demobilizing effect, and no effect. Rather, she finds that overall effect of negativity on voter turnout is negligible, but at the same time, present one condition where demobilizing effect can be observed. However, in order to strengthen her argument for a null effect,

I applied suggestions by Wellek (2010) and Rainey (2014). I chose a size of the parameter that would be interpreted as the substantively meaningful effect, and conducted two one-sided tests. I also performed several robust checks and showed that all of the confidence intervals are within the range that I set.

Over all, reviewing Krupnikov (2011) provided important take away points for my paper in progress. I originally expected to argue for a negligible effect, intending that would still be meaningful breaking the misconceptions people have about dual citizenship. However, I will attempt to find a specific condition in which a significant effect is detected. Moreover, on arguing for a negligible effect, I will also set criteria on the size of the effect that would be substantively meaningful. I will use Two One-Sided Tests and provide 90% confidence interval to evaluate the statistical significance, and perform many robust checks using various methods we learned in class, including permutation tests, bootstrapping, maximum likelihood estimation, and Bayesian approach.

References

- Achen, C. H. (2002). Toward a new political methodology: Microfoundations and ART. *Annual Review of Political Science*, 5(1), 423-450.
- Bagozzi, R. P., Dholakia, U. M., & Basuroy, S. (2003). How effortful decisions get enacted: The motivating role of decision processes, desires, and anticipated emotions. *Journal of Behavioral Decision Making*, 16(4), 273-295.
- Bloemraad, I. (2004). Who claims dual citizenship? The limits of postnationalism, the possibilities of transnationalism, and the persistence of traditional citizenship. *International migration review*, 38(2), 389-426.
- Clark, W. R., & Golder, M. (2006). Rehabilitating Duverger's theory testing the mechanical and strategic modifying effects of electoral laws. *Comparative Political Studies*, 39(6), 679-708.
- Freedman, David A. "Randomization does not justify logistic regression." *Statistical Science* (2008): 237-249.
- Geer, John, Presidential TV Ads 1960-2000
- Geyer, G. A. (2011). *Americans no more*. Garrett County Press.
- Guarnizo, L. E., Portes, A., & Haller, W. (2003). Assimilation and Transnationalism: Determinants of Transnational Political Action among Contemporary Migrants 1. *American journal of sociology*, 108(6), 1211-1248.
- Huntington, S. P. (2004). *Who are we?: The challenges to America's national identity*. Simon and Schuster.
- Kaplan, D. (2009). *Statistical modeling: A fresh approach*.
- Krupnikov, Y. (2011). When does negativity demobilize? Tracing the conditional effect of negative campaigning on voter turnout. *American Journal of Political Science*, 55(4), 797-813.
- Orbell, S., & Sheeran, P. (2000). Motivational and Volitional Processes in Action Initiation: A Field Study of the Role of Implementation Intentions1. *Journal of Applied Social Psychology*, 30(4), 780-797.
- Rainey, C. (2014). Arguing for a negligible effect. *American Journal of Political Science*, 58(4), 1083-1091.
- Rainey, Carlisle, 2013, "Replication data for: Arguing for a Negligible Effect", doi:10.7910/DVN/23818, Harvard Dataverse, V2, UNF:5:R0FapAw5kZ4RSJu+XdGIPA==
- Renshon, S. A. (2000). Dual Nationality+ Multiple Loyalties= One America?. *One America? Political Leadership, National Identity, and the Dilemmas of Diversity*, 232-82.
- Simon, B., & Klandermans, B. (2001). Politicized collective identity: A social psychological analysis. *American psychologist*, 56(4), 319.
- Staton, J. K., Jackson, R. A., & Canache, D. (2007). Dual nationality among Latinos: what are the implications for political connectedness?. *Journal of politics*, 69(2), 470-482.
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. CRC Press.