Review for 'Discrimination that Matters: Replication with Extensions of Perceived Discrimination and Political Behaviour (2020)'

Summary

The replication of the study by Oskooii (2020) about the relationship between perceived discrimination and political behavior is a thorough extension of the original paper. It combines elements of robustness checks and causal inference with a conceptual replication in a different setting. It is an important contribution to the field and allows us to better assess the scope of the original study's claims.

The replication paper gives an overview of the original study, which aims to analyze the impact of societal and political discrimination on "mainstream political participation" and in-group activities. The author relates the paper to other research in the field, building on the original literature review by Oskooii (2020). An extensive description of the replication design follows, dividing it into a robustness check (Study 1) and a conceptual replication (Study 2). Study 1 uses a theoretically motivated altered set of control and predictor variables and a matching procedure that allows for causal inference with observational data. Study 2 extends the original study in two ways, it tests (1) whether the assumed relationships also apply to non-mainstream political participation, in particular support for political violence, and (2) whether the impact on ethnic-based participation can be generalized to samples of immigrants from the first wave of "Causes and Consequences of Socio-Cultural Integration Processes among New Immigrants in Europe" (SCIP) in the UK and the Netherlands.

While the replication paper provides an important extension of the original study, there are some areas in which it should be improved before it can be published: 1) In between many important and correct methodological remarks, some errors were introduced and other concerns were left often. 2) The focus of the paper should be made clearer. 3) Spelling and grammar need to be checked.

Comparability of logistic regression esimates

On p. 7 the author states "Generalised linear models contain unobserved heterogeneity that is independent of the outcome variable and dependent on variances of omitted control variables (Wooldridge 2010; Mood 2010)", which is not correct for all glms (e.g. not for a gaussian glm with identity link). However, the concern about the comparison of coefficients on the link scale and odds ratios can be warranted, but does not necessarily extend to comparisons of differences in probability! (see the same Mood 2010, but also Holm et al. 2015; Breen et al. 2018). If the analyzed units are correctly averaged over in the average treatment effect on the treated, these are comparable as well (cf. Kuha & Mills 2018 about the different meanings of unobserved heterogeneity). In particular, the scaling and comparison issue has nothing to do with maximum likelihood estimation (p. 26).

Due to a misunderstanding of the extent of incomparability, the paper also refrains from interpreting the substantial impact of discrimination, which is falling behind the state of the art for logistic regression models using predicted differences, and (average) marginal or partial effects. Also correctly averaged over odds ratios and risk ratios can be compared. This issue is compounded by interpreting the sign ("directionality") of the results, even when the confidence intervals extend to seemingly similar negative plausible values of the coefficients in the population (e.g., p. 16). A more careful formulation might be better: "The effect is still positive in the observed data".

Control variables, matching and estimands

The original paper and the replication use various alternative versions of what the author calls the "control matrix", discussing them purely as potential confounders while they also change the comparison group in logistic regression (and some matching strategies) and could also be mediating and moderating factors (or outcomes of the political behavior), leading to an increased danger of overcontrolling and collider bias (Spector & Brannick 2011). Why should political discrimination be controlled by social discrimination and vice versa? Is integration in social networks an independent predictor and not an outcome of identity formation? It's also conceivable that social integration strongly moderates the effect of discrimination, making an additive model doubtful. These types of questions should be clarified in the paper.

In general, it would be helpful to clarify the theoretical and empirical estimands (Lundberg et al. 2021) and if they differ between the original study and the replication. This also could be used to contrast how covariates are used in the regression model and matching and how this could support the argument for using the average treatment effect on the treated instead of the average treatment effect (p. 11).

The different matching procedures (optimal full, nearest neighbor, and optimal pair) are not clearly described in the paper, leaving the reader without any potential explanation for the

different results of these matching schemes. If there are good arguments to prefer any one of these matching algorithms, the other ones could also be relegated to the appendix.

Missing assumption tests

Alongside complex robustness checks, I am missing tests of basic assumptions of the generalized linear models used (Pregibon 1981, see Dunn & Smyth 2018: 297ff.), like (most importantly) linearity on the logits and the impact of influential outliers, and less pressingly homoscedasticity (but see Leamer 2010; King & Roberts 2015) and independence of errors. One approach requiring relatively little effort would be plotting the quantized residuals on the main predictors of interest.

While basic assumption tests are generally important to better evaluate the validity of the original study, doing other robustness tests makes missing them more striking. This is particularly important when we take the highly skewed distribution of observed perceived discrimination into account (p. 10), which does not increase my confidence in the linearity of the effect. Especially the difference between the high percentage of people who do not perceive to be discriminated against and people who have a score of 1 is unlikely to be the same as between people on the highest level of discrimination. The very low number of people with very high values in perceived discrimination also makes a comparison of predicted probabilities between maximum and minimum highly uncertain and somewhat dubious. The matching procedure uses only a binary treatment indicator, which implies a different theoretical estimand that should be made more explicit in the paper.

Minor methodological concerns

Like the original study, the paper confuses the terms likelihood (of the data given the model) with probability (of events in a stochastic process, e.g. voting). In addition, hypotheses 3 and 4 are phrased as if in-group attachment and engagement would have been measured on a continuous scale, while they are modeled as binary indicators too. If the main interest is in the unobserved latent variable, it should be measured directly (Kuha & Mills 2018).

The paper currently includes references to tables in the appendix, which are not part of the manuscript. They should either be added in the text (where they are important), to the appendix of the manuscript or referenced as online supplementary material with a fixed identifier instead. In the spirit of the webinar, it would also be helpful to have a short clarification if the original study could be reproduced, accordingly pointing to supplementary material.

Non-significant effects should not be interpreted as no effect (e.g., p. 19 & p. 29) without equivalence tests if the effect is with high confidence below a specific threshold.

Two additional minor issues should be resolved as well: It is unclear which confidence intervals are reported, e.g. in Table 1 (probably 95%-CIs?) and the number of decimal points in tables

should not go beyond the uncertainty of the estimates (e.g. at most 2, maybe just 1 decimal point in most cases).

Lack of focus

The paper discusses arguments supporting effects that were not consistently observed in the replication (p. 29), which is somewhat confusing. They can either be deleted, used in the original derivation of the hypotheses, or more clearly be connected with a discussion of why they might not be observable with the methods used in the replication.

While the author understands his matching analysis as a robustness test, the procedure allows causal inference, which was not the focus of the original paper. Hence, applying matching is more than a simple robustness check and could be much more prominently discussed and expanded upon. These more in-depth causal analyses would probably lead to a longer paper, which could be remedied by a second recommendation: The replication paper currently lacks a clear focus, it is not entirely clear which question it tries to answer, analyzing many interesting aspects at the same time. Focusing either on (1) robustness, (2) causal inference, or (3) generalizability outside the original outcomes and sample could alleviate this. An orientation towards which goal the replication has, could be helpful here (e.g. generalization with different methods and data, Freese & Peterson 2018). Additional material could be provided in the appendix.

Alternatively, the separate aspects could be combined by a central thread discussing if the original paper's claims have causal validity. The current elements could be understood as tests of various threats to causal inference, also allowing for a final discussion of some limitations that have been ignored so far (potential endogeneity, moderating and mediation factors, and especially the danger of misattribution of cause and effect in the cross-sectional analysis). This reorganization would probably necessitate also using matching analysis to analyze the Dutch data-which is currently noticeably missing.

Language and writing

The paper currently suffers from missing parts of sentences, grammatical errors, and other orthographic inaccuracies. Additional editorializing would greatly improve the quality of the paper and might also help with some smaller issues of unclearness. A non-exhaustive list:

- Missing words: "… that provides information about experience of specific types of discrimination (i.e. being)" (p. 5)
- Missing articles: "Survey also did not cover country of birth in equivalent way …" (p. 13)
- Missing brackets: "(being the main form of mainstream political engagement in the original paper and if effects of discrimination found among minorities travel to other similar communities, such as immigrants." (p. 1f.)

- Unclear sentences: “Reminds of the importance of contextual factors in the data collection of the research being replicated, therefore, I am trying to minimise the contextual differences by analysing data on immigrants collected around the time when the original data were collected in the UK.” (p. 12)
- Mixing up singular and plural: “The theory surveyed did not indicate these variables as important controls for in-group attachment.” (p. 14)

Literature

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2018. "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models." Annual Review of Sociology 44(1): 39–54. doi:10.1146/annurev-soc-073117-041429.

Dunn, Peter K., and Gordon K. Smyth. 2018. Generalized Linear Models With Examples in R. New York, NY: Springer New York. doi:10.1007/978-1-4419-0118-7.

Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." Annual Review of Sociology 43: 147–65. https://doi.org/10.1146/annurev-soc-060116-053450.

Gomila, Robin. 2021. "Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis." Journal of Experimental Psychology: General 150: 700–709. doi:10.1037/xge0000920.

Holm, Anders, Mette Ejrnæs, and Kristian Karlson. 2015. "Comparing Linear Probability Model Coefficients across Groups." Quality & Quantity 49(5): 1823–34. doi:10.1007/s11135-014-0057-0.

King, Gary, and Margaret E. Roberts. 2015. "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do about It." Political Analysis 23(2): 159–79. doi:10.1093/pan/mpu015.

Kuha, Jouni, and Colin Mills. 2018. "On Group Comparisons With Logistic Regression Models." Sociological Methods & Research: 0049124117747306. doi:10.1177/0049124117747306.

Leamer, Edward E. 2010. "Tantalus on the Road to Asymptopia." Journal of Economic Perspectives 24(2): 31–46. doi:10.1257/jep.24.2.31.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." American Sociological Review 86(3): 532–65. doi:10.1177/00031224211004187.

Mood, C. 2010. "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It." European Sociological Review 26(1): 67–82.doi:10.1093/esr/jcp006.

Oskooii, K. A. R. (2020). Perceived Discrimination and Political Behavior. British Journal of Political Science, 50(3), 867–892. https://doi.org/10.1017/S0007123418000133

Pregibon, Daryl. 1981. "Logistic Regression Diagnostics." The Annals of Statistics 9(4): 705–24. doi:10.1214/aos/1176345513.

Spector, Paul E., and Michael T. Brannick. 2011. "Methodological Urban Legends: The Misuse of Statistical Control Variables." Organizational Research Methods 14(2): 287–305. doi:10.1177/1094428110369842.