Authors: Frederik K. Kjøller and Lene Holm Pedersen
Title: The Gendered Cost of Politics
British Journal of Political Science
April 9, 2023

All analyses were conducted in R. Please reach out to fkl@ifs.ku.dk if you have any questions.

To comply with GDPR and make responses non-traceable to specific candidates, the data set we share here is without the following columns: name, email, municipality, party. For the figures, where these variables are used, we have created separate, aggregated datasets.

Execute
You can execute the code by creating a .RProj file in the folder that contains the data and code and run the files in this R Project environment or by setting the working directory manually using setwd()

Data files

- data_genderedcost_wide.RData – wide dataset, where each row comprise a political candidates response to the entire survey. includes all complete and incomplete responses.
- data_genderedcost_long.RData. – long dataset used for most analyses where each row represents a job choice. Also contains variables that identify choice situations with dominant options and inattentive choices.
- data_genderedcost_background.Rdata – is a dataset of background variables (which are all present in the above mentioned datasets).
- data_rerun_agg.rdata – is a dataset used for Figure 10. It is aggregated data from the registries in Statistics Denmark which contains data on the rerunning rates for all candidates as well as incumbents in Danish local politics from 1997-2017.
- data_genderedcost_attrition_mun.rdata – is used in Appendices B & F and is a dataset of survey completion rates across municipalities. It is aggregated from candidate-level data, as explained above, because we do not include the municipality column in candidate-level dataset in order to make answers untraceable to individual candidates.
- data_genderedcost_email_party.rdata – is used in Appendix B and is a dataset showing the share of emails across parties, that we had at hand for our data collection.
- data_rep_list.Rdata – is used in Appendix D and is a list of four data sets with aggregated data of each of four

background variables for the entire candidate pool acquired from Statistics Denmark's registries. This is used to benchmark the sample's representativeness compared to the entire candidate pool.

Replication files:

0_cleaning_n_pivoting.R - R script that transforms genderedcost_wide.Rdata into genderedcost_long.Rdata which is a long dataset where each row represents a job choice made by a candidate and is the primary dataset of analysis. The script also creates variables that express whether a job choice entailed a dominant option and whether the respondent is defined as inattentive (by failing to select the dominant option). Further, the script creates a dataset of background variables, like gender, education, political experience, etc., where each row is a respondent. This is merged onto genderedcost_long.Rdata to create background variables for the analyses of job choices, but is also saved as an independent dataset named genderedcost_background.RData, which is used for figure1 and figure2, as well as different figures in the appendices.

1_fig1.R - R script producing figure 1 of victim distributions among men, women and all candidates.

2_fig2.R - R script producing figure 2 of gender differences in assess harassment risk.

3_fig3.R - R script producing figure 3 with marginal means for job attributes of men and women candidates including estimates of the group differences.

4_fig4.R - R script producing figure 4 with marginal means for job attributes of victim and non-victim candidates including estimates of the group differences.

5_fig5.R - R script producing figure 5 with marginal means for the interaction between working environment and institutional position.

6_fig6.R - R script producing figure 6 with estimates for willingness to pay and willingness to work.

7_fig7.R - R script producing figure 7 with marginal means when including and excluding dominant choices.

8_fig8.R - R script producing figure 8 with marginal means for subgroups across gender and victimhood.

9_fig9.R - R script producing figure 9 with marginal means for subgroups political experience.

10_fig10.R - R script producing figure 10 with time series candidates and incumbents tendency to rerun.

appendix_B.R - R script producing all figures in Appendix B.

appendix_C.R - R script producing all figures in Appendix C.

appendix_D.R - R script producing all figures in Appendix D.

appendix_E.R - R script producing the robustness checks in Appendix E.

appendix_F.R - R script producing the robustness checks in Appendix F.

appendix_G.R - R script producing the robustness checks in Appendix G.

appendix_H.R - R script producing the robustness checks in Appendix H.

appendix_I.R - R script producing the robustness checks in Appendix I.

appendix_J.R - R script producing the robustness checks in Appendix J.

appendix_K.R - R script producing the robustness checks in Appendix K.