Revisiting a Natural Experiment: Do Legislators With Daughters Vote More Liberally on Women's Issues?*

By Donald P. Green, Oliver Hyman-Metzger, Gaurav Sood, Michelle A. Zee

Abstract: An intriguing natural experiment arises from the fact that legislators are quasi-randomly assigned some combination of sons or daughters. The pioneering work of Washington (2008) shows that legislators with daughters cast more liberal roll call votes on women's issues. Costa et al. (2019) find that this pattern subsides in more recent congresses and speculate that increasing party polarization might diminish the "daughters effect." The present paper delves more deeply into patterns of change over time by looking at eight congresses prior to the four studied by Washington (2008) as well as eight subsequent congresses, including three not included in Costa et al. (2019). Contrary to the party polarization hypothesis, we find no daughters effect leading up to the period that Washington studied and no effect thereafter.

I. Introduction

One of the most widely-studied natural experiments arises from the fact that public officials are often parents and therefore are quasi-randomly assigned some mixture of boys and girls. Several studies have investigated how legislators' roll call votes are influenced by the sex composition of their biological offspring, and more than a dozen others have investigated the extent to which daughters affect the political and social attitudes of the general public. This literature is summarized in detail in Supporting Information (SI) Table 1.1.

The literature on daughters effects has three recurrent themes. The first is that the applications cover a sprawling assortment of institutions, regions, and historical periods. The second is that studies that report the results of a novel application often find statistically significant results, at least for a subgroup (e.g., fathers whose first child is female). Third, the direction and magnitude of these results vary from one application to the next. When daughters are found to have a liberalizing effect, the explanation is that having daughters impels parents to "protect their daughters from possible gender-based discrimination" (Glynn and Sen, 2015, p. 41), to learn about the challenges of sex-based discrimination, or to accede to pro-feminist pressures from within the household. When daughters are found to have a conservative effect, the explanation is that they "increase conservative views of teen sex" (Conley and Rauscher, 2013, p. 704). Perhaps multiple mechanisms are at work, in which case the daughters effect may be contingent on who is making choices and under what conditions.

Conjectures about context-dependence play an especially interesting role in recent studies of the U.S. House of Representatives. Pathbreaking work by Washington (2008) shows that legislators with daughters cast significantly more liberal roll call votes on women's issues during the 105th through 108th Congresses. Using roll call voting scores compiled by the American Association of University Women (AAUW) for each Congress, Washington (2008) shows that, conditional on the number of children that each legislator has, legislators with more girls are more likely to "vote liberally, particularly on reproductive rights issues" (Washington, 2008, p. 311).

* This project was reviewed and approved by the Institutional Review Board at Columbia University, protocol AAAT-8055. The replication materials are posted at: https://github.com/soodoku/daughters

Reanalyses of Washington's data have affirmed her conclusions (Iacus, King and Porro, 2011; Van Effenterre, 2020). However, an out-of-sample replication study conducted by Costa et al. (2019), which applied similar methods to Congresses 110 through 114, found little evidence of a daughters effect during this period. This null finding also holds for subgroups defined by legislators' gender and party.

What accounts for the discrepancy? Costa et al. (2019) speculate that "the effect of fathering daughters on elite behavior within an institution may be conditional on the intensity of polarization within that setting" (p. 473). This hypothesis is both substantively and methodologically intriguing insofar as a well-identified causal effect is used to assess a much-discussed over time change in American politics, just as repeated audit experiments have been used to assess changes in racial discrimination (Quillian et al., 2017) and repeated investigation of sibling sex composition has been used to study female labor supply in cross-national context (Aaronson et al., 2021).

This paper aims to shed further light on the daughters hypothesis by adducing three new pieces of evidence. We gather data prior to the congresses studied by Washington (2008), starting in 1981. Although partisanship was certainly evident during this period, the degree of party polarization was relatively tame by contemporary standards (McCarty, Poole and Rosenthal, 2016). Second, we gather data after the congressional sessions studied by Costa et al. (2019), so that the resulting dataset spans a total of 20 congresses and affords more precise statistical estimates. Finally, we assess how results change when we track a particular cohort of representatives over time. Specifically, we track the daughters effect over time among legislators who appeared in the Washington (2008) analysis, both in earlier and subsequent congresses. Doing so addresses the possibility that Washington's results were due to the specific roll call votes that appeared on the agenda during the four congresses she studied.

The results show a striking pattern. Like Costa et al. (2019), we find little evidence of a daughters effect in the sessions after those studied in Washington (2008), but we also find no effect in earlier sessions. Contrary to the party polarization hypothesis, the daughters effect was weak during this earlier period of relative party comity. Further, tracking cohorts of legislators over time, we find little temporal variation in the daughters effect over the stretch of 20 congresses. The cohort studied by Washington (2008) displays unusually strong daughters effects; average effects among a broader selection of legislators are close to zero. The concluding section of this paper reflects on the importance of conducting out-of-sample replications of natural experiments.

II. Daughters as a Natural Experiment

The maintained assumption underlying this literature is that the sex composition of a legislator's offspring is determined randomly, conditional on the number of children a legislator has.[1] What are the implications of this randomization for research design and analysis?

First, members of Congress (MCs) without biological offspring are not part of the experiment and should be excluded from the pool of experimental subjects. Second, when measuring the sex composition of legislators' children, researchers using an identification strategy that presupposes random sex selection often exclude adopted children and stepchildren on the grounds that their sex at birth may influence MCs' decision to adopt or remarry. For example, Glynn and Sen (2015, p. 39) exclude non-biological children, noting that "Parents often have strong preferences about a child's gender, and, if given a choice, may opt for adopting a girl over a boy, a boy over a girl, or one child of each gender." Similarly, Conley and Rauscher (2013, p.702) warn that "While the sex of biological offspring may be random, the sex of adopted or even stepchildren is most certainly not." They express con-

[1]Scholars have questioned whether the sex composition of offspring is entirely random. These arguments are reviewed in SI 4.

cern about the daughters-as-natural-experiment literature because "several studies include nonbiological children (adopted and stepchildren), which could bias results because the sex of nonbiological children is not random." (p.700) In this application, the measurement challenge is to obtain reliable information about the sex of each child and whether each child is adopted or a stepchild.[2] Critics of survey-based studies of daughters effects have pointed out flaws in the ways that biological offspring are measured (Conley and Rauscher, 2013; Glynn and Sen, 2015; Hopcroft, 2016), and Washington's data include some cases of adopted children and stepchildren. We set out to rectify these coding issues. Fortunately, it appears that the decision to exclude non-biological children is inconsequential. As Table SI 9.1 shows, the results are effectively unchanged when we analyze the period studied by Washington including non-biological offspring. Third, researchers must measure the total number of relevant offspring as a blocking variable, because even if each child's sex were determined by coin flip, the proportion of girls might vary with the total number of children if parents have preferences over the number of boys and girls.

Like Washington (2008) and Costa et al. (2019), we embarked on a vast data collection effort to determine the sex of each legislator's children. As explained in SI 5, we made use of data collected by previous authors but coded everything freshly using additional archival and online sources. To track down missing information, we supplemented this coding effort by visiting congressional offices on Capitol Hill.

The primary dependent variable is roll call voting on what Washington (p.313) describes as "bills regarding women's issues." Washington (2008) and Costa et al. (2019) measure it using ratings compiled by the American Association of University Women (AAUW). The AAUW rates MCs based on how they vote on select bills. As many have pointed out, the bills that come to a roll call vote in Congress depend on the composition of the Congress, and the votes that are scored by an interest group reflect its policy priorities at that time. As noted in SI 5, the roll call votes that the AAUW used in its scores include issues ranging from reproductive rights to broad economic policies that are thought to help women. Second, few roll call votes are scored in each Congress. For the 115th Congress, for example, ratings of House members are based on six roll call votes. The resulting index therefore has lower reliability than other ideological ratings, such as NOMINATE (Poole and Rosenthal, 1985). In addition, as detailed in the SI, the AAUW data contain some coding errors and ambiguities; in an effort to come up with a more reliable index of women's issue roll call votes, we freshly coded all of the votes based on original legislative voting records. Third, it is unclear whether AAUW scores are materially different from broader and more extensive measures of liberalism-conservatism, such as NOMINATE (Poole and Rosenthal, 1985). The correlation between AAUW and NOMINATE scores averages 0.92 during the period covered by the Washington (2008) study. We therefore consider whether and how the estimated daughters effects change when we use NOMINATE ratings as the outcome variable.

III. Estimation

The causal estimand is the average difference between two sets of potential outcomes among MCs with a given number of children: one set comprises the roll call votes that they would have cast had their children all been male, and the other set comprises the roll call votes they would have cast had they each had at least one daughter. This causal framework can be expanded to consider the "dosage" effects of adding one or more daughters.

If we suppose that legislators who choose to have a given number of children are randomly assigned some proportion of girls, the implied estimator expresses roll call scores as a function of G_i, the number of girls, controlling for C_i, the total number of children. Washington (2008) uses a flexible specification that includes indicator variables for each

[2] A further challenge is to determine the age of each child, since legislators may have additional children while serving in office. Fortunately, given the difficulty of measuring these over-time changes reliably for all MCs, such events seem to be rare, as Washington (2008) points out on p.313.

value of C_i. Washington codes G_i as an integer, although the results do not change appreciably if we explore Costa et al.'s alternative coding schemes, such as scoring G_i as 1 if a member has at least one daughter and 0 otherwise or scoring G_i as the proportion of offspring who are daughters. (Results from an assortment of alternative models may be found in SI 8).

Our main departure from previous work concerns the use of party as a covariate. If it were indeed the case that daughters cause potential political candidates to become more pro-feminist, their party affiliation when elected would be considered a post-treatment variable. MCs with daughters might be more likely to run as Democrats, or they might be more appealing to Democratic voters. For this reason, we are reluctant to control for party, except as a robustness check (See Table SI 8.5). We do, however, follow previous authors by including an indicator for whether the legislator is female, as this variable seems less susceptible to post-treatment bias.

To summarize, our regression model is as follows:

$$(1)\quad \mathrm{Y}_{ij} = \beta \mathrm{G}_i + \gamma_1 \mathrm{C}_{1i} + \gamma_2 \mathrm{C}_{2i} + ... + \gamma_k \mathrm{C}_{ki} + \alpha \mathrm{F}_i + \omega_1 \mathrm{S}_{1j} + \omega_2 \mathrm{S}_{2j} + ... + \omega_s \mathrm{S}_{sj} + \epsilon_{ij},$$

where Y_{ij} are AAUW scores for legislator i in congress j, β represents the average treatment effect of adding a daughter, the C_{ki} indicators mark the number of children a legislator has, F_i indicates whether the legislator is female, and the S_j variables are indicators for each congress. The disturbance term ϵ_{ij} is assumed to be clustered for each legislator; we use the wild bootstrap procedure (Djogbenou, MacKinnon and Nielsen, 2019) to calculate standard errors.

IV. Results

We start by replicating the main results obtained by Washington (2008) and Costa et al. (2019) for the congresses that they studied (see SI 3). Using our data instead of theirs changes neither of their conclusions. Results for the four congresses studied by Washington show that the number of daughters meaningfully increases AAUW roll call scores. Our estimates of the average daughters effect are slightly larger than the ones reported in the original article. The apparent effect using our data is positive and statistically significant ($p = 0.009$). Substantively, the point estimate of 0.056 for the marginal effect of an additional daughter is large enough to be politically consequential. By way of reference, a 0.649 point mean difference divides Democrats from Republicans, and a 0.218 point mean difference divides men from women.

The estimated daughters effect for the five Congresses studied by Costa et al. (2019), on the other hand, is 0.026, which is less than half of what we obtain when analyzing the congresses studied by Washington (2008). SI 3 presents side-by-side comparisons of our estimates and the corresponding estimates using replication data deposited at the time of publication.

A. Pooling 20 Congresses

Table IV.A reports the results from all 20 Congresses combined. If party polarization dampens the daughters effect, going backward in time should increase the apparent average treatment effect. This prediction is not borne out. Instead, the apparent daughters effect (0.019) is smaller than the estimate obtained when we reanalyzed the congresses studied by Costa et al. (2019). The 95% confidence interval ranges from -0.010 to 0.049.

Controlling for party does not materially change the size of the apparent daughters effect. Pooling across all congressional sessions slightly reduces the point estimate from 0.019 to 0.014 (see Table SI 8.5). However, because party is so predictive of outcomes, controlling for party reduces the variance of the point estimates considerably. As a result, some of the estimated daughters effects for specific sessions, notably those immediately before or after the period that Washington studied, appear to be statistically significant at conventional 0.05 levels. The estimates are both insignificant and close to zero for the six earliest congresses and the three most recent ones.

Using NOMINATE scores instead of AAUW scores as a dependent variable produces a similar pattern of results (see SI 8.2). The estimated effect of daughters is close to zero, again with a relatively narrow confidence interval. Interestingly, we also find statistically insignificant daughter effects when using the regression model to predict each member's party (see Table SI 8.4). The lack of relationship between daughters and either liberalism-conservatism or party suggests that the daughters effect is sufficiently subtle that it bears no apparent relationship to two strong correlates of feminism.[3]

B. *Trends over time*

Is there evidence of a monotonic decline in treatment effects over time, in keeping with the hypothesis of increased partisan polarization? Table IV.A, which presents results for each Congress, suggests not. The daughters effect is weakly negative or close to zero during the congresses leading up to Washington's investigation. The only era during which the daughters effect is positive is the one that Washington happened to study.

Is the drift in parameter estimates due to cohort replacement or changes in how a given set of legislators vote over time? To shed light on this question, we split the observations into two subsets: members who served during the sessions that Washington studied and everyone else. As 1 shows, the differences are stark. The cohort that Washington studied displays positive daughters effects over all 20 Congresses. MCs not in this cohort display weakly negative effects throughout. Although the magnitude of the estimated daughters effect varies over time for both cohorts, this temporal variation does not exceed what one would expect by chance. The within-cohort trends do not suggest a change in voting patterns associated with the rise of polarization.

V. Discussion

The daughters effect found by Washington (2008) is a thought-provoking empirical result that seems to demonstrate that roll call votes are influenced by legislators' personal circumstances and experiences. The lack of such effects reported by Costa et al. (2019) are also theoretically suggestive, pointing to a possible shift in legislative decision-making as partisan fealty gains the upper hand on legislators' personal preferences, such as those that might be shaped by their family environments.

The findings presented here seem to support a more mundane interpretation: daughters do not seem to have any appreciable effect on legislators' roll call votes. The findings initially presented by Washington (2008) are statistically persuasive when viewed in isolation, but when viewed in conjunction with data from both earlier and subsequent sessions, it appears that they are something of an outlier. The cohort of legislators that Washington happened to study do exhibit the behavior she ascribes to them, but other cohorts exhibit no such pattern, and the average effect now appears to be close to zero.

This replication failure does not appear to be due to "researcher degrees of freedom" (Simmons, Nelson and Simonsohn, 2011) because the findings reported in Washington (2008) are robust to an assortment of measurement and estimation choices. Changing how we measure family composition, which covariates we adjust for, or which roll call votes we tally scarcely affects the results. Washington's discovery seems to have resulted from statistical variation in voting patterns across congressional cohorts.

[3]This null finding also has an important methodological implication. Ordinarily, discerning the effects of daughters on roll call votes would be complicated by the fact that the liberalizing effects of daughters could affect candidate recruitment and electability. This concern about what amounts to post-treatment attrition subsides if daughters truly have no effect.

Table 1—Effect of the Number of Daughters on AAUW Score Controlling for Indicator Variables for the Number of Children and MC's gender

	Dependent variable: AAUW																				
	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Daughters	−0.006	−0.004	−0.022	0.002	0.002	0.010	0.021	0.033	0.077***	0.040*	0.068***	0.046**	0.019	0.046**	0.002	0.041	0.030	0.006	−0.004	−0.010	0.019
	(0.018)	(0.020)	(0.022)	(0.026)	(0.023)	(0.027)	(0.025)	(0.027)	(0.025)	(0.024)	(0.026)	(0.021)	(0.019)	(0.021)	(0.025)	(0.026)	(0.021)	(0.020)	(0.029)	(0.026)	(−0.010, 0.049)
N	362	366	361	361	364	368	377	386	397	393	392	392	390	400	392	400	401	396	395	377	7,670
Adj. R^2	0.012	0.029	0.012	0.008	0.012	0.015	0.102	0.055	0.094	0.090	0.112	0.055	0.054	0.066	0.048	0.054	0.075	0.103	0.113	0.178	0.065

Note: *p<0.1; **p<0.05; ***p<0.01. The coefficients for indicator variables for the number of children and MC's gender are suppressed. The pooled model includes fixed effects for each Congress.

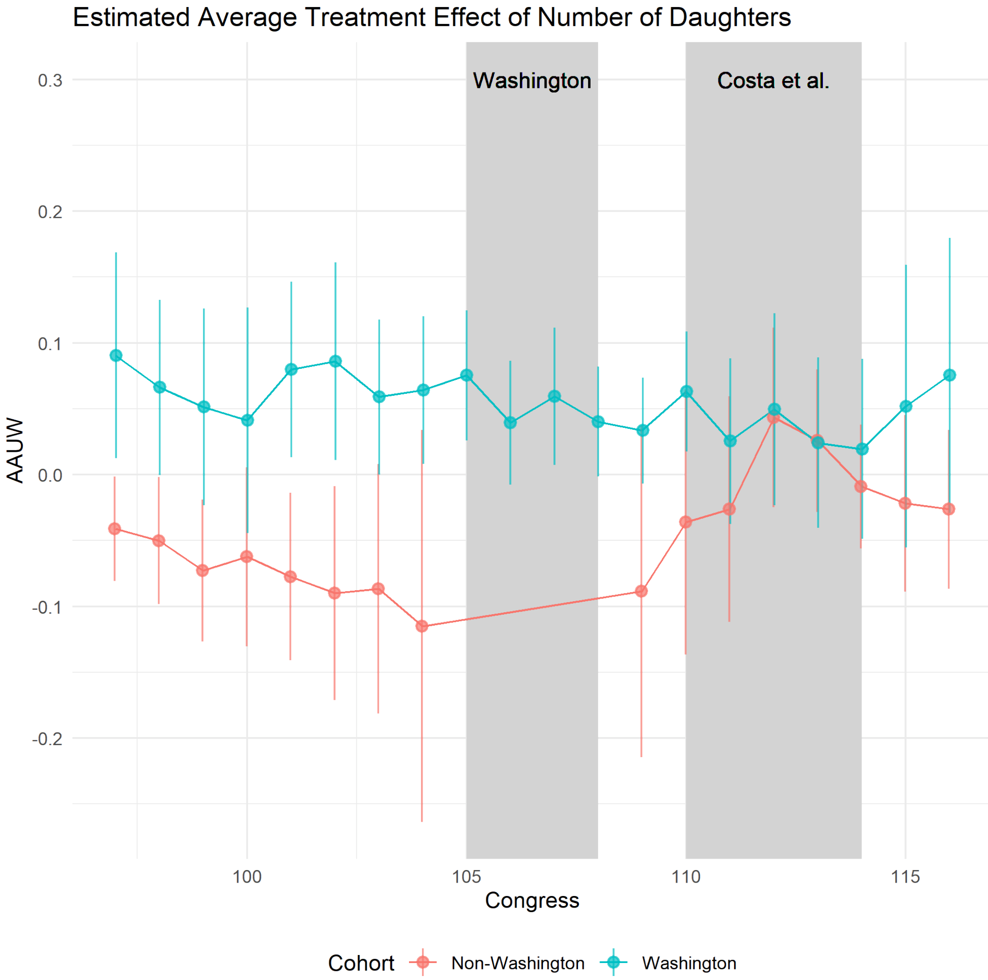

Figure 1. Tracking the Daughters Effect Over Time for the Cohort of Legislators Analyzed by Washington (2008) and for All Other Members

To the extent that something systematic underlies the gap between the initial and subsequent results, it may be a variant of the file-drawer problem: natural experiments that generate noteworthy findings receive attention, while those that do not are consigned to oblivion. In the context of daughters effects, the number of historical eras, countries, and institutions provides a large set of potential draws from the sampling distribution.[4] This interpretation has testable empirical implications: natural experiments, especially those that produce theoretically engaging results, should have sub-par performance when subjected to out-of-sample replications.[5]

[4]All the more so when authors zero in on specific subgroups that seem to manifest effects, a relatively common occurrence according to Table SI 1.1.

[5]For a recent example on the subject of reservations for women government officials, see Clayton, de Kadt and Dumas (2022).

REFERENCES

Aaronson, Daniel, Rajeev Dehejia, Andrew Jordan, Cristian Pop-Eleches, Cyrus Samii, and Karl Schulze. 2021. "The effect of fertility on mothers' labor supply over the last two centuries." *The Economic Journal*, 131(633): 1–32.

Clayton, Amanda, Daniel de Kadt, and Natasha Dumas. 2022. "Daughters Do Not Affect Political Beliefs in a New Democracy." *Journal of Experimental Political Science*, 1–11.

Conley, Dalton, and Emily Rauscher. 2013. "The effect of daughters on partisanship and social attitudes toward women." *Sociological Forum*, 28: 700–718.

Costa, Mia, Jill S Greenlee, Tatishe Nteta, Jesse H Rhodes, and Elizabeth A Sharrow. 2019. "Family ties? The limits of fathering daughters on congressional behavior." *American Politics Research*, 47(3): 471–493.

Djogbenou, Antoine A, James G MacKinnon, and Morten Ørregaard Nielsen. 2019. "Asymptotic theory and wild bootstrap inference with clustered errors." *Journal of Econometrics*, 212(2): 393–412.

Glynn, Adam N, and Maya Sen. 2015. "Identifying judicial empathy: does having daughters cause judges to rule for women's issues?" *American Journal of Political Science*, 59(1): 37–54.

Hopcroft, Rosemary L. 2016. "Comment on Lee and Conley (2016),"Does the Gender of Offspring Affect Parental Political Orientation?": Lee and Conley's Null Findings Are Likely due to Measurement Error in their Measure of Sex of First Child." *Social Forces*, 95(2): 893–897.

Iacus, Stefano M, Gary King, and Giuseppe Porro. 2011. "Multivariate matching methods that are monotonic imbalance bounding." *Journal of the American Statistical Association*, 106(493): 345–361.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches.* MIT Press.

Poole, Keith T, and Howard Rosenthal. 1985. "A spatial model for legislative roll call analysis." *American journal of political science*, 29(2): 357–384.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H Midtbøen. 2017. "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time." *Proceedings of the National Academy of Sciences*, 114(41): 10870–10875.

Simmons, Joseph P, Leif D Nelson, and Uri Simonsohn. 2011. "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological science*, 22(11): 1359–1366.

Van Effenterre, Clémentine. 2020. "Papa does preach: Daughters and polarization of attitudes toward abortion." *Journal of Economic Behavior & Organization*, 179: 188–201.

Washington, Ebonya L. 2008. "Female socialization: how daughters affect their legislator fathers." *American Economic Review*, 98(1): 311–32.