Appellate Court Assignments as a Natural Experiment:

Gender Panel Effects in Sex-Discrimination Cases

Robert S. Erikson
Department of Political Science
Columbia University
rse14@columbia.edu

Word count 8,493

Paper prepared for 11th Annual NYU-CESS Experimental Political Science Conference February 9th & 10th 2018

This paper argues that causal inference about panel effects on the US Courts of Appeals can be advanced by taking advantage of the random assignment of judges to cases. The substantive example is gender panel effects in sex-discrimination cases. The paper makes an empirical claim that on sex-discrimination cases, judge assignment can be considered as-if random. Treating the data as a series of natural experiments, the paper confirms that the presence of a female judge on a panel influences male co-panelists to vote liberal. It also shows how exploiting random case assignment facilitates testing for a causal mechanism—here a combination of endogeneity effects (conformity) and context (females as better advocates).

[*] I thank Chris Blattman, Alexander Coppock, Sandy Gordon, Don Green, John Kastellec, Anja Kilbarda, Jeff Lax, Kelly Rader, Laura Stoker, Greg Wawro, Dane Worley, Joon Yang, and Miranda Yaver for their helpful comments. I thank Joon Yang and Anja Kilbarda for valuable research assistance. Special thanks are due Christina Boyd and Andrew Martin for their guidance with understanding the Boyd *et al.* sex-discrimination data.

Appellate Court Assignments as a Natural Experiment: Gender Panel Effects in Sex-Discrimination Cases

I. Introduction

One of the most compelling findings in the subfield of judicial politics is the discovery of "panel effects" on the 12 circuits of the US Courts of Appeals. Cases before these courts are typically tried by three-judge panels whose members are chosen randomly (in theory) from the circuit's roster of available judges. These three-judge panels are distinguished by their unanimity--with upwards of ninety percent of written decisions decided 3-0—a pattern that can be only partially explained by judges sharing a common interpretation of legal and constitutional issues. Also contributing to the frequent unanimity is a norm of consensus, whereby judges tend to acquiesce to the views of their co-panelists. This conformity is inferred from studies of panel effects, most notably of how judges respond to their panel's party composition. Judges are more likely to take the liberal side when serving with Democratic rather than Republican co-panelists (Sunstein, et al., 2006; Kastellec, 2011; Epstein, *et al*. 2013 Fischman, 2015).

Panel effects takes on special importance when gender or racial diversity is involved. Female judges who hear gender discrimination cases and African-American judges who hear racial discrimination cases may hold special sway over their co-panelists. More might be involved in these cases than the usual conformity, with males giving unique deference to females on gender-based discrimination cases (Farhang and Wawro, 2004; Peresie, 2005; Boyd, et al., 2010; Fischman, 2015) and white judges giving special deference to racial minorities on race-based discrimination cases (Farhang and Wawro, 2004; Kastellec, 2013). In short, diversity on

Appellate Courts might affect case outcomes more than the simple addition of a few new voices would indicate.

Absent from this literature, however, is any attempt to advance causal inference by leveraging the random assignment of judges to analyze the data as a natural experiment. In the ideal natural experiment, treatment and control conditions are assigned randomly by a real-world mechanism without the intervention of a researcher (Dunning, 2012). In other judicial venues, scholars have exploited random assignment of judges to link judges' characteristics to sentencing policies and evaluate their downstream effects (Kling, 2006; Green and Winik, 2010; Aizer and Doyle, 2015).[1] Why not do the same for panel effects on the US Courts of Appeals?

The literature on panel effects typically derives causal claims from multivariate equations predicting judges' votes from their co-panelists' demographic characteristics, sometimes along with fixed effects for circuits and time. Some recent studies (Boyd et al., 2010; Kastellec, 2013) introduce propensity score matching to the task. With propensity score matching, treated judges are matched with controls based on similarity in the multivariate propensity score predicting the likelihood of treatment (e.g., serving with female or nonwhite co-panelists) While propensity-score matching has become a popular tool for causal inference, its use is restricted to observational studies as a hopeful substitute for random assignment. When

[1] Beyond the judicial realm, political science examples of natural experiments include analyses of the random assignment of ballot order (Ho and Imai, 2008), legislative committee memberships (Broockman and Butler, 2015), legislative terms (Titiunik, 2016), and military draft eligibility (Erikson and Stoker, 2011).

treatment and control conditions are randomly assigned, propensity-score matching adds no value for the obvious reason that randomly assigned treatment and control cases cannot be predicted.

Existing non-experimental studies of panel effects pool data across circuits and years, so that judges are compared with other judges in other circuits at different times. The frequency with which Democrats, African-Americans, or female judges populate a circuit's judge pool varies by circuit and time, possibly related to variation in culture or docket. And the differences between Democrats, blacks, or females within one circuit and Republicans, whites, or males in a different circuit at a different time can differ from the relevant comparisons within specific circuits at the moment of random selection.[2] Thus, as commonly recognized by analysts (e.g., Boyd *et al.* 2010, fn. 13, p. 394); random assignment offers no cloak of protection when comparing cases across circuits and at different time points.

The causal effect of interest is the difference in two potential outcomes--the votes by two "subject" judges conditional on whether the third judge on the panel is a Democrat/African-American/female or the counterfactual, a Republican/white/male. To estimate this panel effect, the ideal experiment would compare subject judges' votes when administered the random draw of a "treatment" judge—a Democrat/ black/female—versus the random draw of a "control" judge—Republican/white/male from the pool of available judges in the circuit at the particular

[2] As an analogy, in the US states the differences between Democrats and Republicans within states are different from the differences across states, e.g., Massachusetts Republicans versus Louisiana Democrats.

time. The causal analysis would seemingly be straightforward, perhaps as a simple difference of means between subject judge votes under treatment and control conditions.

Despite the theoretical advantage of random assignment, researchers have had their reasons to be wary about whether this experimental ideal could be approximated with US Courts of Appeals data. Studies have challenged whether supposedly random assignment of judges to cases is sufficiently random in practice and this challenge must be taken seriously. Further, one could doubt whether a sufficient number of matched treatment and control cases could be found for a meaningful analysis. And it is not obvious what the ideal estimation strategy should be, considering that the unique feature of federal appellate courts are three-judge panels where all three judges can influence each other.

The present paper argues for treating votes on the US Courts of Appeals as being drawn from a series of natural experiments. The substantive focus is the question of gender panel effects in Title VII sex-discrimination cases—whether male judges are more likely to take the liberal position when a female judge is assigned to their panel. The data are from Boyd *et al*.'s (2010) study of gender panel effects as they pertain to sex-discrimination cases. Relying on propensity score matching, Boyd *et al.* find their strongest evidence of gender panel effects in sex-discrimination case---one instance where theory suggests gender influence would be likely.

Where feasible, random assignment trumps matching for estimating causal effects. For sex-discrimination cases, the paper shows that assignments of judges to cases can safely be treated as-if random, i.e., without confounding. The sites are the 12 circuits of the US Courts of Appeals. The time band is the two-year interval between federal elections. Within a specific circuit and time band, the presumably random draw of a female or male panelist (of the same party) is unrelated to other variables affecting the outcome of the case.

Exploiting random case assignment as a series of natural experiments offers several advantages. First, limiting comparisons to randomly assigned cases within circuit-periods dissolves the threat to causal inference from unmeasured confounders and assures common support. Second, it offers statistical simplicity and transparency with no complicated multivariate analysis necessary. Third, as shown below, the experimental framework allows advances in understanding the causal mechanisms of judges' influence on each other.

This paper proceeds as follows. Section II describes the research design. Section III evaluates the degree of randomness in the assignment of judges to cases. Section IV presents the evidence for a gender effect—the average difference in the judges' votes (conditional on party) contingent on the lottery's draw of a male or female judge. Section V estimates the gender *panel* effect, the effect of a female panelist on male judges' votes. Section VI shows how random assignment can advance understanding of the causal mechanisms that drive a gender panel effect. Section VII discusses the possibility of a within-subject design to look for downstream effects of female judges' votes on their male co-panelists' votes in future cases. Section VII concludes.

II.. Research Design and Data

Even if Appellate Court judges are assigned to panels on a random basis (see more on this below), they are distributed inefficiently from the analyst's perspective. Treatments (female co-panelist) are assigned at a lesser rate than the control condition (all-male panel). Further, the treatment of female judges is distributed unequally across circuits and across time. As a potential complication, a male control judge on one panel could be a subject judge on a different panel. And the two subject judges influenced by the treatment or control judge influence each others' decisions. These facets do not deter from the value of observing the effect of a random draw of a female onto a panel rather than having all males. But they do influence the specific contours of the research design.

The cases here are drawn from the 415 sex-discrimination cases from the Sunstein *et al*. data base of Appellate Court decisions over multiple issue areas, 1995-2002—the exact cases examined by Boyd *et al.* Due to small sample size, the 19 rare cases with two female judges are set aside, yielding 396 cases and 1188 judge votes. A female judge is empaneled on precisely one-third of the 396 cases; two-thirds of the panels are all-male.

Each solo female judge on a panel is a potential treatment to be paired with one or more controls, *i.e*., male judges within the same circuit within the same two-year period, where the treatment and control judges are of the same political party. The subjects are the two male judges administered either the treatment (female co-panelist) or control (all male panel) condition. Comparing female treatment judges with male control judges of the same party ensures that any difference across treatment and control outcomes are not due to party affiliation. Further, female judges are paired with their male counterparts (of the same party) only when the partisanship of the two subject panelists is identical for treatment and control—both Democrats, both Republicans, or split. Given the strong influence of partisanship on Appellate Court panels, this step enhances the precision of the estimates.[3] Thus, for instance, a female

[3] Party of the judge is defined by appointing president's party. Below, further controls are introduced for judicial ideology via Epstein *et al*.'s (2007) Judicial Common Space (JCS) scores

Republican judge with two male Democratic co-panelists is exactly matched with male Republican control judges with two Democratic male co-panelists. Not all of the 132 cases with one female judge can be matched with an all-male counterpart for the same circuit-period. However, 95 panels with one female judges can be compared to contemporaneous all-male panels cases with the identical partisan division.[4]

We are interested in the influence of the female treatment judges and their male controls over their panels' two subject judges. A complication is that the two subject judges also interact with each other. This requires that the dependent variable comprises not the votes of individual subjects (the typical approach to analyzing Circuit Court panels) but rather the votes of pairs of subjects serving on the same panel. The votes of the two male subject panelists are treated as the mean of the pair rather than two separate observations. Given judges' influence on each other, votes by different subject judges on the same panel should not be considered to be independent events.

The dependent variable is the difference between subject panelists' votes in the treatment case with a female judge and subject panelists' votes in the matched cases with a male judge.

---

and Bonica and Sen's (2017) DIME scores. JCS scores are derived from Common Space score of the appointing president or (when appropriate) the home-state senator. DIME scores derive from the Common Space scores of the judge's political contributors.

[4] On average. female treatment judges are matched with 4.4 males. When a female judge is matched with *N* matched male controls, the controls are each weighted 1/*N*.

In the language of causal inference, this is an estimate of the average treatment effect (ATE).[5] This is *matching with replacement* (Abadie and Imbens, 2006); a male judge can serve as a control judges for more than one female judge.

Treatment and control cases are drawn for the circuit within two-year periods. Over two years, the pool of available judges will inevitably show some fluidity. However, entries and exits within two-year intervals are minimal. Ninety-seven percent of the votes by active judges and senior judges in the Sunstein *et al.* sex-discrimination data set were by judges serving in the judge pool throughout the two-year period.

III. Random Assignment and Sex-Discrimination Cases

Before proceeding, we must confront the question of whether the assumption of random case assignment is sufficiently fulfilled in sex-discrimination cases. In theory, cases are assigned randomly to prevent litigants from judge-shopping and judges from lobbying for assignments. In practices the exact procedures are imperfectly understood. Sometimes randomization is assured by the use of a lottery device, but there are many exceptions. For instance, chief judges or court administrators might strive to ensure equal workloads, fixed rotations, or the scheduling convenience of judges (Hall, 2010, Levin, 2017).

Empirical examinations have found assignment patterns that are not likely from pure random assignment. Hall (2010) finds that the partisan makeup of panels varies by subject

[5] The estimated average effect on the two subject panelists incorporates both a direct effect and an indirect effect involving the two subject panelists' influence on each other. For instance, treatment/control Judge A influences subject Judge B who influences subject Judge C and similarly influences subject Judge C who then influences Judge B.

matter (within circuit and term) than expected by chance. For sex-discrimination cases, however, Hall finds no partisan bias. Chilton and Levy (2015) estimate the degree to which the partisan divisions (0, 1, 2, or 3 Republicans) depart from what would be expected by chance given the partisan makeup of available judges in the circuit. They discover four circuits (2, 8, 9, DC) with patterns that deviate from chance by at least the .10 level of significance, and a disturbingly low .003 *p*-value when data is pooled across all 12 circuits. However, this finding is inappropriate for our purposes here. For the many instances when three judges are empaneled together for multiple days rather than only one, Chilton and Levy count every day's panel meeting separately, even though the repeat panels (about one-third of the cases) are not independent events. Chilton and Levy conduct a secondary analysis in which they count multiple-day panels only once. By this more appropriate test, only two circuits show sub-.10 *p*-values. Pooled over the full set of 6,675 separate panels, the pattern of party composition is significant at a mere .36 level (Chilton and Levy, p. 47).

Rather than require knife-edge certainty about the purity of random assignment as a strict lottery, let us accept the more modest standard that panel assignments sometimes take into account considerations that depart from pure random assignment but that these considerations are unrelated to the facts of the case. In the most thorough investigation of Circuit Court procedures, Levy (2017) finds no indication of panel composition designed to tilt case outcomes: "The factors considered by the courts ranged from those related to logistics to those touching on collegiality, but none had an ideological bent" (p. 104). In the language of potential outcomes, the crucial question is not strict randomness of assignment but strict ignorability—the absence of confounding influences. Conveniently, this assumption is subject to empirical testing,

Selection of Female Judges Related to Male Judge Ideology?

One potential confounder could be if selection of a female panelist somehow triggers the selection of male co-panelists based on ideological disposition. The evidence argues strongly against such a problem. With circuit-period controlled, the presence of a female panelist is statistically unrelated to male panelists' partisanship, Epstein *et al*.'s (2007) Common Space ideology scores, Bonica and Sen's (2017) DIME scores, or minority racial status.[6]

Are Female Assignments Related to Case Facts?

A more worrisome confounder would be if females are assigned to panels based on the case's anticipated outcome. Could it be that, despite judicial norms, chief judges or court administrators systematically place female judges on those sex-discrimination cases where it is hard (easy) to reach a liberal verdict? If so, the effect of case difficulty could masquerade as the influence of female colleagues on male judges.[7]

It is possible to test for an effect of the strength of the female plaintiff's case on the gender balance of the panel. The Sunstein *et. al* sex-discrimination data contains 19 binary facts about each case, which can serve as instruments for case outcome. First, we predict the case

[6] Males who serve with females are very slightly more nonwhite and liberal than males on all-male panels within their circuit and period, with non-significant *p*-values for party (.34), Common Space scores (.16), DIME scores (.66), and non-white status (.44).

[7] Apart from violation of random assignment, a potential selection bias could arise from litigants' being motivated to settle prior to trial upon learning the panel's gender composition. See Boyd et al. (2010), fn. 35, p. 406. However, during the period of this analysis only the DC circuit publicized the panel composition prior to the hearing date. According to Jordan (2007) the DC circuit experiment led to no excess frequency of settlements.

outcome (liberal or conservative) from these 19 variables (or a subset) to create a first-stage index of likelihood of a liberal outcome; i.e., deciding for the female plaintiff. To ensure comparisons within circuit for two-year periods, fixed effects for circuit-period are imposed. Then scores from the prediction equation serves as instruments to predict the assignment of one or more female judges. Predictability in this second-stage would be a sign of non-randomness.

Table 1 presents three separate first-stage equations, each in logit format to accommodate the binary nature of the dependent variable. Model 1 includes all 19 fact variables; they are collectively statistically significant. However, four appear to matter for the appellate outcome: the direction of the lower court verdict, unequal pay as an issue, significant damages are involved, and the absence of First Amendment issues. The remaining 11 variables are collectively insignificant (p=.18). Model 2 incorporates only the four statistically significant variables from Model 1. Model 3 includes only the lower court verdict---the most salient case fact when panels are chosen. Appellate Courts tend to defer to the originating court's decision. As the lone outcome predictor in the equation, the lower court verdict is quite statistically significant; female plaintiffs win only 29 percent of the time when appealing a lower court verdict, but 51 percent of the time when successful in the lower court.

In the second-stage analysis, can the predicted values from these linear equations predict the likelihood of a female panelist? As shown at the bottom of Table 1 the answer is negative. The linear equation predictions from the three first-stage models are all statistically insignificant as predictors of female panel participation and even conflict on the sign of the supposed relationship.[8] Whether a female is represented on a panel appears to be unrelated to expectations of whether supporting the female plaintiff would be easy or hard.[9]

The most compelling evidence of *de facto* randomness is a direct comparison of the 95 treated cases for the analysis which follows and their controls in terms of the probability of a pro-plaintiff opinion. The treated cases (with a female judge) are virtually identical to their controls (all-male) in their estimated probability of a pro-female verdict based on case facts. For each model, the mean differences are slim and far from statistically significant. In effect, treatment and controls are almost exactly balanced in terms of the probability of a liberal verdict.

Random Assignment? A Summary

By tradition, Circuit Court judges are assigned "randomly" to three-judge panels. Although assignments are not always made via a randomization device such as a lottery wheel, tradition and current accounts suggest that case facts are unrelated to judge assignments. The analysis here reveals no reason to suspect otherwise for sex-discrimination cases. For the following analysis, assignment of judges to panels is treated as-if random.

[8]There are many alternative variations to the equations of Table 1: logit vs. OLS, all cases vs. excluding cases with more than one female, and instrumenting via the equation prediction vs. the predicted probability of a liberal vote. These alternative setups yield similar findings.

[9] A simpler test is to estimate the reduced form equation predicting case outcomes directly from the 19 fact variables (plus circuit-periods). Their collective *p*-value is a non-significant at 0.18.

Table 1. Probable Case Outcome Based on Case Facts and the Selection of Female Panelists

First stage logit equation: Predicting a verdict favoring female plaintiff from case facts	(1)	(2)	(3)
Lower Court supports plaintiff	0.64 ((0.31)	0.64 (0.30)	0.98 (0.27)
Claim unequal pay	0.74 (0.37)	0.70(0.35)	
Sued under First Amendment	-2.00 (0.97)	-1.40 (0.84)	
Damages are a major point of contention	0.98 (0.41)	1.01 (0.40)	
Pregnancy discrimination	0.90 (0.53)		
Complains employer acted in retaliation	0.52 (0.28)		
All plaintiffs female	0.05 (0.36)		
Title IX Claim	1.28 (1.03)		
Section 1983 Claim	0.96 (0.59)		
Constructive discharge from employment	0.81 (0.44)		
Procedural issues dominate	0.39 (0.44)		
Plaintiff suing under state law	-0.18 (0.32)		
Claim an illegal denied promotion	-0.55 (0.37)		
Claim illegally not hired	0.36 (0.51)		
Claim illegally fired	-0.21 (0.30)		
Sued under Section 14A	0.36 (0.69)		
Section 1981 claim	0.13 (0.44)		
Age discrimination claim	0.16 (0.38)		
Emotional distress claim	0.15 (0.50)		
Control for Circuit-Period?	Yes	Yes	Yes
P-value, fact covariates	.006	.0001	.0003
N	410	410	410
Stage 2: Logit equation predicting female panelist(s) from propensity to support female plaintiffs (predictions from Stage 1)			
Control for Circuit-Period?	Yes	Yes	Yes
Logit equation predicting plaintiff success	0.17 (0.14)	0.14 (0.19)	-0.39 (0.29)
P-value, logit equation predicting plaintiff success	.221	.460	.182
N	402	402	402
Mean probability of a decision favoring the female plaintiff (from Stage I equation). *N*= 95 paired treatments and controls.			
Treatment Cases (One Female Panelist)	.371	.366	.352
Control Cases (No Female Panelists)	.347	.348	.356
Difference	.024	.018	-.004
p-value	.31	.33	.73

Standard errors in parenthesis. The Stage 1 dependent variable is a liberal (pro-plaintiff) verdict (1) or conservative (pro-defendant) verdict (0). The Stage 2 dependent variable is presence (1) or absence (0) of at one least female on the panel. The key Stage 2 independent variable is the linear equation prediction from the Stage 1 equation. Of the 415 cases, the equation *N*'s exclude those without variation within the circuit-period in the verdict (stages 1,2) or panel gender (stage 2).

IV. Gender Differences in Judging Sex-Discrimination Cases

If female judges influence their male colleagues to vote for the liberal position on sex-discrimination cases, a primary assumption is that female judges are more likely to cast liberal votes than males. Unless that were true, the search for a liberalizing effect of female judges on male panel-members would make little sense. Thus, our necessary precursor to estimating the gender panel effect is to estimate the effect of gender on voting on sex-discrimination cases.

The idea of "gender effects" might suggest a search for residual gender differences in a dependent variable after controlling for many variables. But that type of search would be contrary to the potential outcomes model of causal inference (Rubin, 2005) and is not the direction here. The causal question in this study is the difference in a judge's likely vote, contingent on whether the randomly selected judge is a male or a female. For litigants, this is a relevant causal question. More specifically, we are interested in gender differences in voting given the party affiliation of the judge plus the partisanship of the two male subject judges on the panel. For instance, suppose we know that the judge to be selected will be a Democrat and that the other two judges are one male Democrat and one male Republican. How much does the random selection of a male versus a female Democrat affect the probability of a liberal vote?

The first set of columns in Table 2 present the test for gender effects—are female judges more liberal than their male counterparts on sex-discrimination cases? The 95 observations are the cases decided by one female judge, each matched with one or more male judges. The outcome of interest is the difference between the female judge's decision and that of the male controls. The table shows that female judges—regardless of their party or the parties of their co-panelists—are about 15 percentage points more likely to vote liberal in sex-discrimination cases than their male counterparts. With the 95 differences of means pooled together, the gender difference is significant at the .01 level (in calculating the significance level, the standard errors are conservatively clustered by individual judge; the 95 female judge decisions are made by only 37 separate female judges). Clearly, female judges are more liberal than their male counterparts, and to a greater degree than the ten percent differential reported by Boyd *et al.*

V. Gender Panel Effects

We turn to our central question: does a female judge's presence on a panel induce her male co-panelists to vote more liberally? As in the previous section, female judges are matched with control male judges from the same circuit and period who are also members of her party. Further, treatment and control cases are also matched on the partisanship of the subject judges, either both Democrats, both Republicans, or split. The dependent variable is the difference (treatment versus control) in the mean votes of the two sets of subject panelists.

The second set of columns in Table 2 shows the results. With an average treatment-control differential of 13.7 percentage points, the gender *panel* effect is almost as large as the average differential between drawing a female versus male judge (14.6). The estimate varies little with the partisanship of the treatment judge or the treated panelists. The net result is in the center of the range of "12 to 14 percent" reported by Boyd et al. (p. 406).

Figure 1 depicts the gender panel effect, this time in terms of the combined response of the two subject judges. The figure makes clear that subject panelists rarely splitting their votes. They almost always decide the same way—as a result of a shared understanding of the facts, influencing each other, , and the influence of the third (treatment or control) panelist.

Table 2. Judge's Gender and Voting on Appellate Court Panels									
		Gender Effect (Votes of Randomly Selected Female Treatment Judges and their Male Controls)				Gender Panel Effect (Effect of Treatment/Control Judge Gender on Subject Male Judges)			
Party of Treatment/ Control Judge	Party Division Of Subject Judges	*N*	Female Treatment Judge (% Vote Liberal)	Male Control Judge (% Vote Liberal)	Difference (F – M) (% Vote Liberal)	*N*	Given Female Treatment (% Vote Liberal)	Given Male Control (% Vote Liberal)	Difference (F – M) (% Vote Liberal)
Republican	2 Reps.	17	35.3	18.3	+16.6	17	27.0	18.3	+8.7
	Divided	13	46.3	27.9	+18.4	13	42.3	30.0	+`12.3
	2 Dems.	4	50.0	37.2	*	4	50.0	37.5	*
	All	34			+17.0	34			+10.3
Democrat	2 Reps.	31	48.4	34.1	+14.3	31	43.5	28.2	+15.3
	Divided	28	46.4	31.5	+14.9	28	44.6	25.8	+18.8
	2 Dems.	2	25.0	75.0	*	2	50.0	75.0	*
	All	61			+13.3	61			+15.6
All Cases		95	45.3	30.7	+14.6	95	41.0	27.3	+13.7
				Std. Err.	(5.3)			Std. Err.	(5.7)
				p-value	.01			*p*-value	.03
The 95 cases represent 95 female treatment judges, exactly matched with male controls of the same party with the same partisanship for subject judges, within the same circuit-period. Standard errors are clustered by female treatment judge.									

Thanks to random assignment, the two distributions in Figure 1 are identical in terms of circuit-periods, the party of the treatment/control judge, and the partisan division of the two subject judges. The sole difference is the gender of the treatment judge. With a female judge treatment, about fourteen percent more of the male subjects vote for the female plaintiff than when under the control condition of an all-male panel.

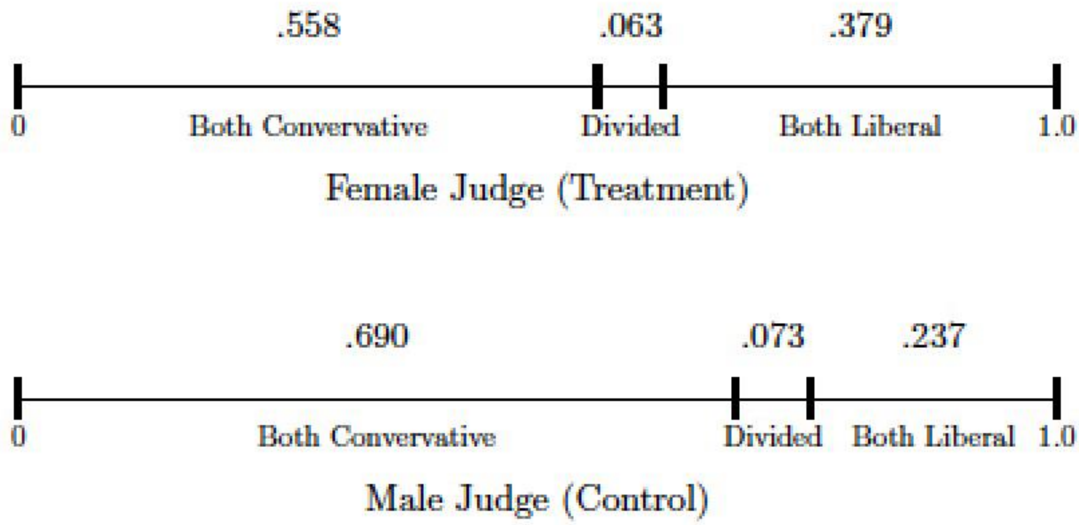

Figure 1. Gender panel effects in sex-discrimination cases. Graphs represent the votes of pairs of male subject panelists. Treatment of a female or male judge is randomly assigned. Subject judges (co-panelists) are all males.

Robustness Checks

The coefficient plot in Figure 2 displays a series of robustness checks from variations in the exact specification. Model 1 repeats the estimated gender panel effect from Table 1, with matching on subject-judge partisanship. Model 2 (unlike those that follow) ignores matching on subject-judge partisanship. The tradeoff is an expanded *N* but a loss of precision from the abandoned partisan control. Model 3 restricts the time band to one year (circuit-years instead of circuit-periods). Model 4 restricts the observations to those where the two subject judges served in the circuit's judge pool for the entire two-year period. Model 5 excludes the four circuits where Chilton and Levy (2015) report suspected non-randomness.

Model 6 expands the matching to incorporate the direction of the lower court verdict, Here the tradeoff is the added precision from the additional control but a lesser *N* due to the tighter matching restrictions. Model 7 controls for the case facts differently, regressing the subjects-treatment control vote differential on the differential in predicted outcome from four case facts (Table 1, Model 2), and observing the constant term. This yields some efficiency gain while retaining all 95 original cases. Model 8 includes the facts differential from Model 7 while adding the Epstein et al (2007) and Bonica and Sen (2017) measures of judge liberalism for a further gain in precision. Reassuringly, all eight estimates are within the narrow 12.6-14.6 range, accompanied by slight but predictable variation in their standard errors.[10]

Random Assignment vs. Propensity Score Matching

How do this paper's estimates of gender panel effects compare to those by Boyd *et al.*? The sizes of the estimated gender panel effects are similar: 13.6 percent vs. Boyd *et al's* 14.1 percent. This comparison is limited, however; while the present analysis separates gender panel effects from party panel effects (e.g., effects of female Democrats vs. male Democrats),

[10] We can also ask about the 19 panels with two females. Fourteen of these female pairs could be matched with similar male panels on circuit-period and the partisanship variables. In pairs of two, female judges voted 8.6 percentage points more liberal than their matched male counterparts. The third (treated) judge votes only 1.8 percentage points more liberal with the treatment of two females rather than two males—as if two female judges are less influential on a third male judge than one female's influence on two male co-panelists. Needless to say, the small sample size requires extreme caution.

Boyd et al.'s analysis does not.[11] The important comparison of the two studies is methodological; whereas we treat the data as-if from a natural experiment, Boyd et al. instead attempt to control for confounders via propensity-score matching. When allowed by random assignment, treating the data as if from a natural experiment is superior to propensity-score (or other) matching designs. Let us review the reasons why this is the case here.

Let us turn the spotlight on Boyd *et al.*'s propensity score equation that for male judges predicts the treatment of serving on a gender-diverse panel. The predictors include JCS score, JCS score squared, racial minority status, and minority status times the JCS score, as well as circuit dummies. Treatment cases are matched to controls (with replacement) based on nearest-neighbor propensity scores plus exact matching on year of decision, a judicial experience dummy, and direction of the lower-court decision.[12] It turns out that more than half of the treatment-control matches are between treatment in one circuit and control judges in a different circuit, thus potentially introducing bias from omitted variables at the circuit level. Further, with circuit and year controlled, the combined judge-level variables in the propensity equation are collectively non-significant (*p*-value=.56) suggesting they have no independent predictive value

---

[11] The gender panel effect can be estimated blind to partisanship using matching with replacement within circuit-years. The estimate is 15.1 percentage points. This result is locally variable, depending on the relationship between the circuit's gender distribution and its partisan distribution.

[12] See Boyd et al., footnote 26. The exact match on judicial experience is found in the code, although not mentioned in the footnote. Each treatment case is matched with a set of six (weighted) controls.

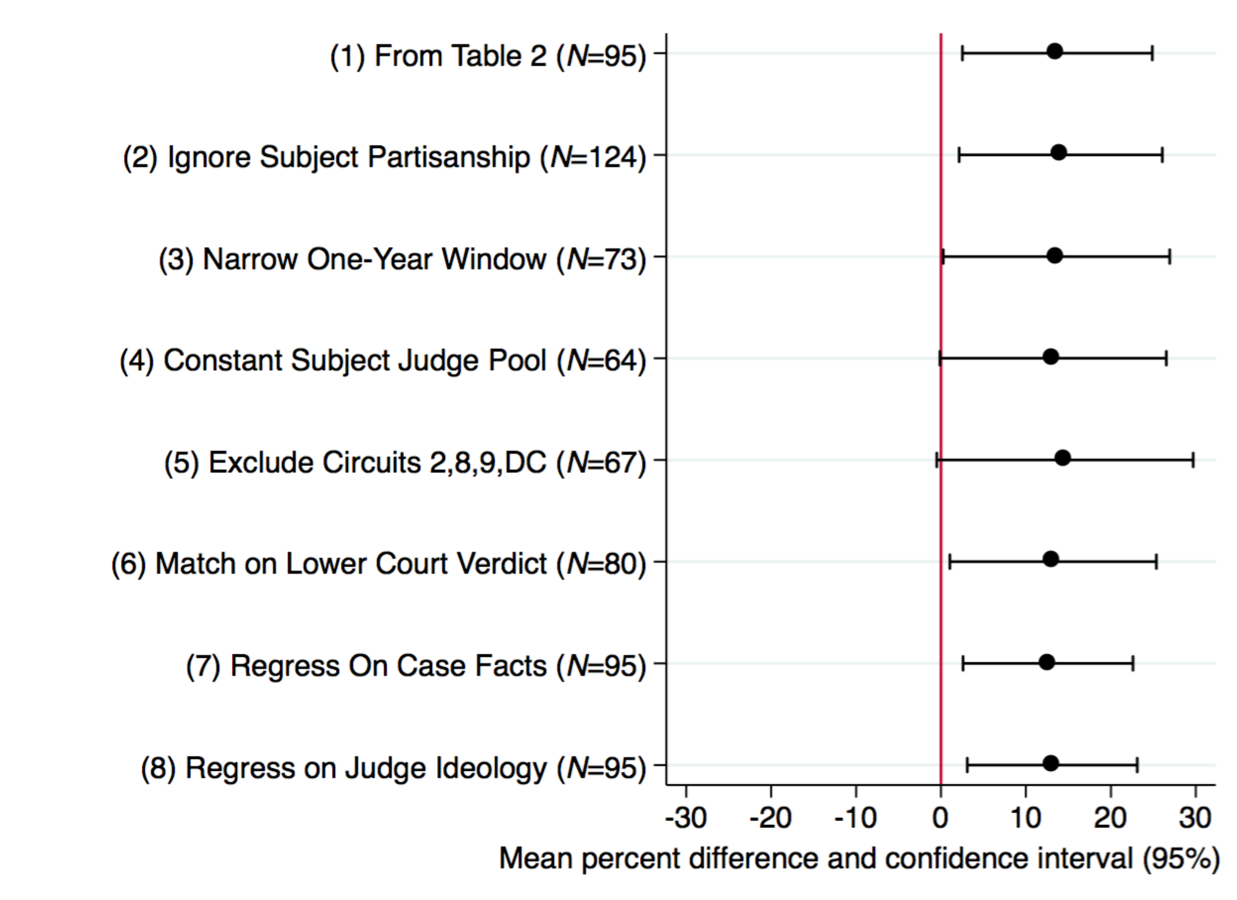

Figure 2. Coefficient Plot of Alternative Specifications for Estimating Gender Panel Effects. Mean differences are between the mean percent liberal of treatment male judges and the mean percent liberal of control male judges where the treatment is a single female judge on the panel and the control condition is an all-male panel. For Models 7 and 8 the "mean difference" is the intercept of the regression equation, when the control variables (treatment-control differences) are at zero. Case facts in Model 7 = the predictions from Table 1, Model 2. Ideology for Model 8 = Epstein et al.'s Judicial Common Space scores and Bonica and Sen's (2015) DIME scores.. The Bonica and Sen measure of ideology is complete for all subject judges (treatment and control conditions) for 77 of the 95 cases. In the remaining 18 cases with missing data, the treatment-control difference is set to zero. Model 8 also incorporates Model 7's case facts. For all models, standard errors are clustered by female treatment judge.

for predicting panel assignment. This of course is as exactly what is expected with random assignment! The net result is that the matched variables add no particular value, while the between-circuit comparisons risks bias that could easily be avoided. But what if our assumption of as-if random assignment is actually in error so that the comparison of treatment and control cases within circuit years is biased by unanticipated confounders? If so, propensity score matching would offer no protection as it would suffer the same bias from within circuit-period unanticipated confounders, while allowing bias from avoidable between-circuit comparisons.

By the simple trick of making comparisons only within circuit-periods, the problem of bias from comparing judges from different circuits (or different time periods) vanishes. If the assumption of as-if random assignment to cases hold, causal inference can proceed with the confidence that arises from analyzing data from natural experiments.

VI. The Causal Mechanism

The panel gender effect (13.7 percent) is 94 percent as large as the gender effect (14.6 percent). Seemingly, when the random selection of a female panelist generates an increase in the probability of a pro-plaintiff vote, it generates a nearly equal increase in probability that each male co-panelist will vote pro-plaintiff. What accounts for this apparently large influence of female judges over their male colleagues? Identifying the causal mechanism is a daunting but important challenge for experimental work (Imai, *et al*., 2011; Bullock and Ha, 2011). As we will see, treating panel effects as if they emerge from a natural experiment can enhance our ability to account for the causal effect of female judges on their panel colleagues' opinions.

To account for the causal mechanism behind panel effects, Fischman (2015) distinguishes between *endogenous* and *contextual* effects. By an endogenous effect, judges react to their fellow panelists' *opinions* which are reflected by their demography. It is the female judges' liberal views and not their gender that influences male co-panelists to vote liberal. Endogenous effects reflect the norm of consensus on US Court of Appeals panels, as judges tend to conform to the views of their colleagues.[13]

By a contextual effect, the demography of the persuading judge also matters, independent of the persuading judge's position. When siding with the female plaintiff in sex-discrimination cases, female judges might be more able to sway their male colleagues than would a male judge advocating the same position (with male judges being perhaps more persuasive when arguing the contrary). A further possibility is an opinion-context interaction effect. An extreme example would be if only female judges are capable of influencing male judges in sex-discrimination cases.

The special curiosity about gender panel effects in sex-discrimination cases is sparked by the possibility that more than the usual consensus effects are operating. When advocating for a female plaintiff, a female judge might make the more compelling argument, perhaps schooling their male colleagues based on their personal experience with discrimination in the workplace (Boyd *et al*., 2011). Moreover, as advocates for female plaintiffs, female judges may be perceived to hold the most intense opinions. Their male co-panelists could see opportunities

[13] The idea of consensual panel effects is treated in a general way here. For nuance, see the discussions in Sunstein et al., 2006, Chapter 4, Kastellec, 2011, and Fischman, 2015.

for implicit logrolls—conceding on issues special to female judges in return for a payoff on other issues down the road.

Is any model more plausible than the other? As a mental experiment, assume an endogeneity effect but no contextual effect. Figure 3 presents the model, adopted from

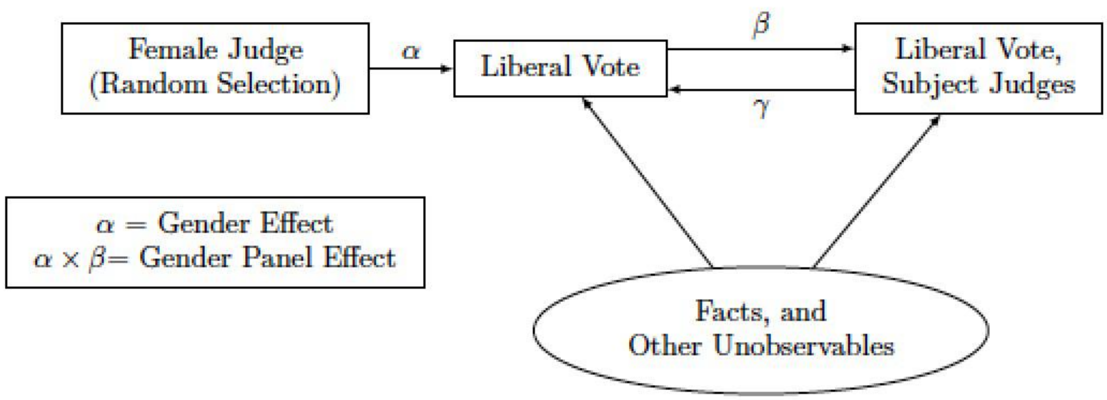

Figure 3. Hypothetical model of gender panel effects as a strictly endogenous effect due directly to opinions and not gender. Subject judges are all male.

Fischman (2015). Assuming it has no direct effect, gender serves as an instrument for estimating judges' influence on their co-panelists via two-stage least squares (TSLS). The estimated effect of any one judge's vote, male or female, on each co-panelist would be the ratio of the 13.7 percentage-point gender panel effect to the 14.6 percentage point gender effect. The result of this exercise is the implication that the votes of any two judges' votes pivot with .94 probability on the vote of the third judge's vote. Implausibly, any judge, male or female, could almost always convince any other judge.[14] The model's implausible result suggest an error in the assumption of no direct "contextual" direct effect of the treatment/subject judge's gender on the subject judges' vote.

Next consider the implausible state of the world under the alternative extreme, whereby only female judges are capable of influencing male judges on sex-discrimination cases. With probability .94, female judges would have the power to sway the votes of their male co-panelists. Meanwhile, male judges would be incapable of influencing each other.

A plausible conjecture is that the truth lies in a middle ground, with both endogenous and context effects; when judges vote liberal, both males and females can influence their co-panelists, with females being better at it (males may be more influential on conservative votes). But does random assignment allow further side evidence to support this interpretation?

Table 4 provides such supporting evidence. The pair of four-cell subtables reanalyzes the 95 treatment cases and their 95 sets of controls as observations of the frequencies by which the vote of judge A predicts the votes of Judges B and C. Judge A is either the female treatment judge or her male control and Judges B and C are the male subject co-panelists from Table 2. The variation in results for treatment female judges and control male judges provides further evidence regarding the causal mechanism.

---

[14] Although implausible. a 0.94 effect would be theoretically possible. It would reflect both direct influence of Judge A on Judges B and C but also the indirect effects A→B→C and A→C→B. Using Fischman's (2015) calculus, the direct effect of one judge's opinion on another would be "only" 0.48, less than the maximum of 0.50.

The "percent differences" can be treated as OLS regression coefficients. While they should not be treated as causal, as-if random assignment allows considerable leverage. The treatment (female judge) and control (male judges) cases share similar distributions of case facts and identical distributions of circuit-periods and identical partisan divisions of treatment/control

Table 4. Subject (Male) Judges' Vote by Treatment (Female) or Control (Male) Judges' Vote

	Male "Control" Judge			Female "Treatment" Judge	
Subject Judge Vote:	Conservative	Liberal		Conservative	Liberal
Percent Liberal	2.2	84.3		2.9	87.2
Percent Conservative	97.8	15.7		97.1	12.8
N	131.7	58.3		104	86
	OLS *b*=0.821			OLS *b*=0.843	

Column percents equal 100. Treatment and Control *N*'s each = 190, two subject judges for each of 95 treatment or control judge, where controls are weighted so that each treatment judge's control judges sum to *N*=1. "Treatment" and "Control" samples are matched to be identical on circuit-period, treatment/control judge party and the partisanship of the two subject judges.

judges and subject judges. Treatment and control cases presumably share the identical covariance between the votes of Judge A and subject Judges B and C that are due to the influence of the case facts. They presumably share similar effects of male Judges B and C on Judge A. They differ only in a possible differential influence of female (treatment) and male (control) judges on subject male judges.

Table 4 readily disconfirms that gender panel effects are solely due to an endogeneity explanation, which would require that judges influence their fellow panelists a whopping 94 percent of the time. The two coefficients—the percentage point differentials of 0.843 percent (female judges) and 0.815 (male controls) as if "effects" on the subject judges' votes fall below 0.94. In fact, as estimates of "effects," these coefficients must be biased upward since they also incorporate both the influence of subject judges on treatment/control judges plus the shared variance from the facts of the case. All factors push the OLS coefficients in the positive direction. This validates the idea that the gender panel effect cannot be attributable solely to endogeneity, that is, that gender panel effects are due solely to the greater liberalism of female judges.

Table 4 also disconfirms the opposite extreme that only female judges have influence. If only female judges could influence male judges on sex-discrimination cases, Table 4's coefficients for males would represent only the influence of case facts on control and subject judges, while those for females would additionally reflect their singular influence on male subjects. Yet the coefficient for female treatment judges is a mere .028 greater than that for male controls.

The evidence from Table 4 suggests a mixed explanation; liberal judges, male or female, have some sway over their male co-panelists to also vote liberal. But liberal decisions by female judges are more persuasive than liberal decisions by males. Random assignment ensures that the range of facts and partisan contexts are similar for the treatment and control cases. A crucial distinction is that as judge A, female judges vote liberal an extra 14.6 percent of the time—the gender effect from Table 2.

Male (control) judges take the liberal side only 31 percent of the time, what we might call the "easy" cases; when they do, their co-panelists agree with them 84 percent of the time (Table 3). Meanwhile, female (treatment) judges take the liberal side a more generous 45 percent of the time, expanding the range of pro-plaintiff votes by almost 40 percent beyond the easy pro-plaintiffs votes by their male controls. It follows that if male and female judges were equally influential with their male co-panelists, liberal-voting female judges would induce the lesser rate of liberal voting due to their expansion of the zone for liberal verdicts to include more difficult cases than their male controls. Yet when female judges vote liberal on this expanded set of easy and more difficult cases, their male co-panelists comply at an 87 percent rate (Table 3), even more than when serving with male control judges and their stricter standards for a liberal vote. This pattern is possible only if on the "easy" cases on which both male controls and female treatment judges side with the plaintiff, the female judges are more persuasive.[15]

We now have a general explanation for why the random selection of a female or male judge has almost as much impact on the votes by male co-panelists as on the votes of the female or male judges themselves. Female judges cast more liberal votes and liberal votes by one judge often influence co-panelists to also vote liberal. Further, under circumstances when either a female or male judge votes liberal, the female judge's vote carries more sway with co-panelists.

[15] Table 3's data is also consistent with male judges being more influential when taking the conservative position, with male subjects slightly more likely to vote conservative when the male control judge votes conservative than when the female treatment judge (with her stricter threshold for a conservative vote) votes conservative. The differential for conservative decisions is of lesser consequence however, since almost all treatment/control conservative votes are accompanied by conservative votes by the two subject judges.

There are not sufficient degrees of freedom to claim more than the existence of both endogenous and contextual effects. However, the reader is invited to consult the appendix, which expands on the possibilities.

VII. Within-Subjects Design: A Before-After Test

When the causal process is simple consensus-seeking, the phenomenon presumably is temporary, with little reason to expect that a panel's consensus in one case will carry over to how the panel's judges decide future cases. The extraordinary effectiveness of female judges when advocating for female plaintiffs in sex-discrimination cases may provide the exception. Arguably, female judges can educate their male colleagues about the perils females face in the workplace in a way that male advocates cannot. This is Boyd et al.'s (2010) *information* effect, as female judges "possess unique and valuable information emanating from shared professional experiences" (p. 391) to offer their male co-panelists. This information effect could persist beyond the case at hand. As exposure to the informed arguments by female judges sensitizes male judges to women's perils in the workplace, their newly learned awareness could affect their downstream decisions in future sex-discrimination cases.

If serving with a female judge increases male judges' empathy for the female position in sex-discrimination cases over the long run, we should see male judges voting more in support of female plaintiffs *after* serving with female liberal advocates than before. This suggests a within-subject design, following judges' decisions over time. One handicap, however, is that we must measure the timing of panels by the dates of their published decision where the ideal would be by the dates of oral arguments, thus possibly blurring the measure of timing. But the major handicap is a lack of statistical power, due to the relative paucity of usable cases.

To appreciate the challenge, consider the most basic test, male judges with a three-case sequence where they vote on all-male panels immediately before and after the middle panel that includes a liberal-voting female judge. There are only 36 such cases. To enhance precision given the small sample size, votes are adjusted as residuals from the predictions of a logistic regression equation accounting for male judges' votes on all-male panels from co-panelists' partisanship plus the probability of a liberal verdict from case facts and circuit-period (Table 1, column 2). By this test, the net gain in liberal voting from the first to the third panel is 7.4 percentage points, which might seem impressive. However, this estimate is accompanied by an 8.2 standard error and a decidedly insignificant *p*-value of only .38.

Many variants of the test for long-term gender-panel effects are possible. Their results vary in magnitude, generally in the hypothesized direction, but rarely approaching statistical significance. In summary, the data yield a modest hint—but are capable of no more—of a longer-term liberalizing effect when male judges are exposed to female panelists.

For perspective, consider the following. Within panels, judges are remarkably pliable in how they vote, bending in the direction of their randomly assigned co-panelists. Yet they also cling to their varying ideological predispositions. It could be unrealistic to expect an opinion on one panel induced by the panel's composition to extend by more than a trace to future panels with new co-panelists and a new sets of factual constraints. Yet the data are at least consistent with the idea that exposure to female co-panelists contributes modestly to a cumulative growth in male judges' empathy for female plaintiffs. The exact answer, however, defies easy detection by the instruments employed here. Hopefully future research can validate the precise long-term impact of panel composition.

VIII. Conclusions

With sex-discrimination cases as data, this study demonstrates how the understanding of panel effects on US Court of Appeals decisions can be enhanced by taking seriously the random assignment of judges to cases. At least for the example of female judges and sex-discrimination cases, US Circuit Court decisions can indeed be treated as-if random from a series of natural experiments. The usual advantages of experimentation for causal inference apply.

In the language of experimental research, the study was conducted as follows. For gender-discrimination cases, pairs of male judges (subjects) were randomly administered a third judge on the three-judge panel who was either a female (treatment) or (with greater frequency) a male (control) of the same party. Comparing judicial votes under treatment and control conditions within the same circuit and two-year period, male judges vote about 14 percentage points more liberal when randomly assigned to serve with a female judge rather than on an all-male panel.

The findings here do not upend findings of earlier studies. The magnitude and reported significance level of the estimated gender panel effects are comparable to evidence from conventional observational analyses based on multivariate analysis of judge votes pooled across circuits and time. From a substantive standpoint, this is reassuring.[16] One might ask then, what is gained by reanalyzing the data as a set of natural experiments?

---

[16] It does not always turn out that experimental results agree with results from earlier non-experimental statistical analysis. For a classic case of disagreement, see Lalonde, 1986.

The central advantage of course is that evidence from random assignment (here within circuit-periods) inspires greater confidence than from techniques such as matching across circuits and periods. Treating the data as a set of natural experiments assuages concerns about the influence of unmeasured variables that could bias estimates of causal effects. Further, the search for multiple control variables or the elaborate matching of cases (across geography and time) becomes unnecessary. Statistical displays can be simple. An additional advantage is the facilitation of sensitivity testing regarding the causal sequence.

The puzzle in this study was that the random selection of a female panelist had almost as large an effect on the votes of male co-panelists as on the vote by the female judge herself. The analysis here suggests that both male and female judges can influence their male co-panelists, with the additional factor that liberal-voting female judges are more persuasive at influencing co-panelist votes than male counterparts who offer similar liberal verdicts. Speculatively, the persuasive talent of female liberal judges might stem from their unique value as informants about discrimination in the workplace plus their greater empathy and involvement in this gender-based issue. Limited evidence suggests the possibility that this influence persists downstream beyond the case at hand.

Admittedly, the argument for treating US Court of Appeals case assignments as-if random hinges on assignments being as-if random, in the sense that a judge's selection is independent of the other panelists selected and the case facts. This assumption appears to be empirically valid for female judges and sex-discrimination cases. The intent of this study is to provide a template for analyzing panel effects on the Appellate Courts as a natural experiment not only for judges' gender and sex-discrimination cases, but also beyond. There are other issue areas, other kinds of panel effects, and other US Appeals Court data sets. Further exploitation of random case assignment on the US Courts of Appeals should enrich understanding about behavior on these important courts.

References

Abadie Alberto and Guido Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica* 74:235–67

Aizer, Anna and Joseph J. Doyle Jr. 2015. "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges." *Quarterly Journal of Economics*. 130: 759-83.

Bonica, Adam and Maya Sen. 2017. "A Common-Space Scaling of the American Judiciary and Legal Profession." *Political Analysis*. 25: 114-121.

Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science*. 54: 389-411.

Broockman, David E., and Daniel M. Butler. 2015. "Do Better Committee Assignments Meaningfully Benefit Legislators? Evidence from a Randomized Experiment in the Arkansas State Legislature. *Journal of Experimental Political Science* 2: 152-163.

Brown, J. Robert, and Allison Herren Lee. 2000. "Neutral Assignment of Judges at the Court of Appeals." *Texas Law Review*. 78: 1037-1117.

Bullock, John G. and Shang E. Ha. 2011. "Mediation Analysis is Harder than It Looks." Chapter 15 in James Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, eds., *Cambridge Handbook of Experimental Political Science*. New York: Cambridge University Press.

Chilton, Adam S. and Marin K. Levy. 2015. "Challenging the Randomness of Panel Assignment in the Federal Court of Appeals." 01 *Cornell Law Review*: 1-56:

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences* New York: Cambridge University Press.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland. 2007. "The Judicial Common Space" *Journal of Law, Economics, and Organization*, 23: 303-325.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*. Cambridge, MA: Harvard University Press.

Erikson, Robert S. and Laura Stoker. 2011. "Caught in the Draft: The Effect of Vietnam Draft Lottery Status on Political Attitudes." *American Political Science Review*, 105: 221-237.

Farhang, Sean, and Gregory Wawro. 2004. "Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making." *Journal of Law, Economics, and Organization*, 20: 299-330.

Fischman, Joshua B. 2015. "Interpreting Circuit Court Voting Patterns: A Social Interaction Framework." *Journal of Law, Economics. And Organization*. 31: 801-832..

Green, Donald P. and Daniel Winik. 2010. "Using Random Judge Assignment to Estimate the Effects of Incarceration and Probation on Recidivism among Drug Offenders." *Criminology*. 48: 357-387.

Hall, Matthew. 2010. "Randomness Reconsidered: Modeling Random Judicial Assignments in the US Court of Appeals." *Journal of Empirical Legal Studies*. 7: 574-89.

Ho, Daniel E., and Kosuke Imai. 2008. "Estimating Causal Effects of Ballot Order from a Randomized Natural Experiment: The California Alphabet Lottery, 1978-2002." *Public Opinion Quarterly* 72: 216-40.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review*, 108: 765-789.

Jordan, Samuel P. 2007. "Early Panel Announcement, Settlement, and Adjudication. *Brigham Young university Law Review*, 55. Available at http://digitalcommons.law.byu.edu./lawreview/vol2007/iss1/2

Kastellec, John. 2011. "Hierarchical and Collegial Politics on the U.S. Courts of Appeals," *Journal of Politics* 2011, 73: 345-61.

Kastellec, John. 2013. "Racial Diversity and Judicial Influence on Appellate Courts," *American Journal of Political Science.* 57:167-83

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review.* 96: 863-876.

Lalonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs," *American Economic Review*, 76: 604-620.

Levy, Marin K. 2017. "Panel Assignment in the Federal Courts of Appeals," 103 *Cornell Law Review* 103:65-116.

Peresie, Jennifer L . 2005. "Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts." *Yale Law Journal* 114(7): 1759-90.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association*. 100: 322-31.

Sunstein, Cass R., David Schkade, Lisa Michelle Ellman, and Andres Sawicki. 2006, *Are Judges Political? An Empirical Analysis of the Federal Judiciary.* Washington: Brookings Institution Press.

Titiunik. Rocio. 2016. "Drawing Your Senator From a Jar: Term Length and Legislative Behavior". *Political Science Research and Methods.* 4: 293-316.

Appendix

Modeling Endogeneity and Demography Effects

This appendix presents simulations of gender panel effect as a function of both the general impulse for conformity (the endogeneity effect) and demography (females are the more influential advocates for liberal positions but less influential when voting conservative). The constraints imposed are Table 2's estimated parameters for the gender effect and the gender panel effect, plus the percentages shown in Table 4. Conditional on these inputs plus some further modest and reasonable assumptions, it is possible to scan for possible models of how the 13.6 percent gender panel effect arises.

Figure A1 graphs the votes of female treatment judges and their male controls as a function of the implied strength of the female defendant's case in the judge's eyes, based on data from Table 2. With random assignment, in expectation the distributions of this unobserved

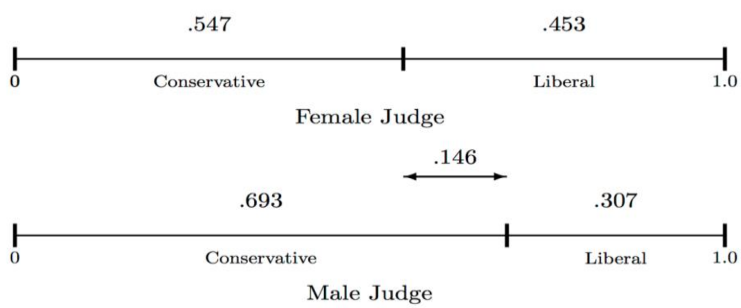

Figure A1. Decisions of Female Treatment Judges and their Male Controls as a function of the support for the female plaintiff's case, in percentiles.

strength variable are identical for female treatment judges and their male controls. Consistent with the discussion in the text, female judges have a lower threshold at which they side with the female plaintiff.

Figure A2 presents some possible models of how the subject judges' votes depend on the treatment/control judge's position on the case and their gender. The variation across the figures stem from the one degree of freedom available—the assumed degree of initial similarity of views between treatment/control judges and subject judges.

The graphs of Figure A2 are based on the following model which models the latent unobserved interval-level judgments of treatment/control judges and subject judges in terms of degree of support for the plaintiff's position.

The latent judgment X_{ik}^{*} of the *i*th treatment/control judges on the *k*th case is:

$$X_{ik}^{*} = \Omega_k + \beta F_i + u_{ik}$$

where $\Omega_k \sim N(\kappa, 1)$ = the shared considerations of case *k*

F_i =0 if the *i*th judge is male and 1 if female

$u_{ik} \sim N(0, \sigma^2)$ = the *i*th judge's residual evaluation of case *k*.

The *i*th treatment/control judge's vote $X_{ik} = \begin{pmatrix} 1 & X_{ik}^{*} > 0 \\ 0 \text{ otherwise} & \end{pmatrix}$

The latent judgment Y_{jk}^{*} of the *j*th subject (male) judge on the *k*th case is:

$$Y_{jk}^{*} = \alpha + \Omega_k + \lambda X_{ik} + \phi F_{ik} + \theta F_i X_{ik} + e_{jk}$$

where $e_{jk} \sim N(0, \sigma^2)$ = the *j*th judge's residual evaluation of case *k*.

The jth treatment/control judge's vote $Y_{jk} = \begin{pmatrix} 1 & Y^*_{jk} > 0 \\ 0 \text{ otherwise} \end{pmatrix}$.

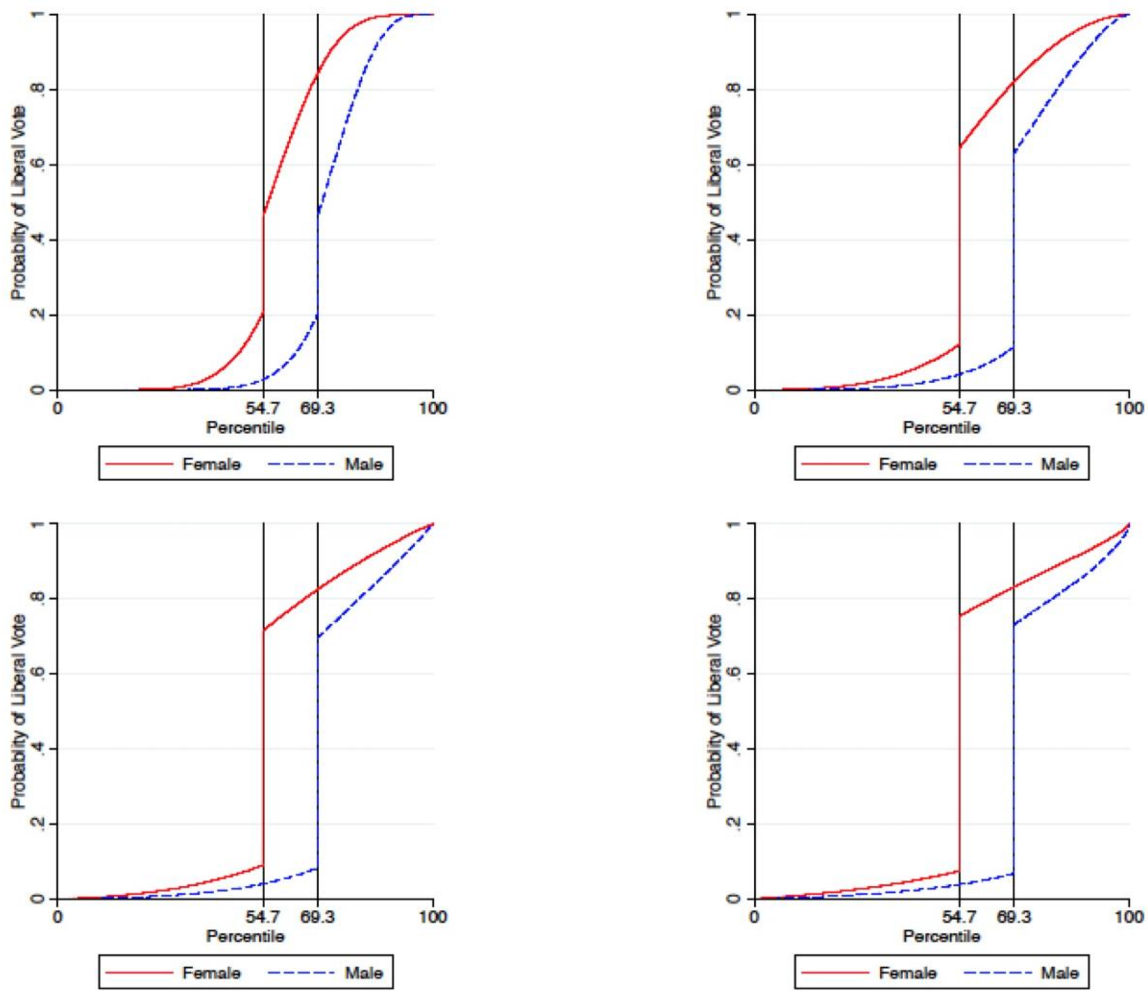

Figure A2. Alternative simulations of the effect of the treatment/control judge's vote and gender plus case facts on the subject judge's probability of a liberal vote. The *x*-axis represents the treatment/control judge's degree of support for the plaintiff's position, as a percentile. Each version is consistent with the findings from Tables 2 and 4. The variations are due to the magnitude of σ^2, the variance of the judges' disturbance term relative to the variance due to common understanding of the facts. As σ^2 increases from the upper-left panel (0.0625) to the lower-right panel(1.00) , the relative impact of the treatment/control judge's opinion matters more and their gender matters less.

Note that $u_{ik}$and e_{jk}share a common variance σ^2.

Figure A2 depicts the proportion mean Y_{jk} as a function of the percentile of X^*_{ik} now scaled as the *percentile* of support for the plaintiff, with the results constrained to be consistent with the results of Tables 2 and 4. The one degree of freedom is the size of σ^2, the variance of the judge-specific variance relative to the variance of the shared evaluation Ω_k.

In the figures, the horizontal axis represents the percentile preferences of the treatment judge, female (treatment) or control (male). The two thresholds for the treatment judge's acceptance of a liberal verdict are set at the 54.7 (females) and 69.3 (male) percentiles. At each threshold the effect of a female or male judge's decision on their co-panelists' probability of voting liberal is a discontinuity in the probability curve. The net space between the two curves represents the 14.6 percent effect of a female treatment rather than the all-male control condition.

The discontinuities at the thresholds represent the difference it makes on co-panelists at the cut-point between the treatment or control judge voting liberal or conservative. The sizes of these hypothetical effects are conditional on the assumed degree of similarity between the initial inclinations of treated and treatment judges. Yet in each graph, a good share of the 13.7 percent gender panel effect is due to the lesser influence of male control judges than their female counterparts when they cast liberal votes. That is, at the higher percentiles in terms of support for the female plaintiff—where both male controls and female treatment judges vote for the plaintiff—the female judges enjoy a higher rate of influence over their male treated co-panelists.

Each graph in Figure A2 shows a visible differential between the influence of male and female treatment/control judges, conditional on the percentile of support for the plaintiff. Given the constraints of this exercise it is not possible to present a graph reflecting the estimated gender and gender panel effects based on the endogeneity effect alone. Thus, while the treatment/control judges' votes matter in terms of their influence on co-panelists' votes, so does their gender.

Also of interest, the gender effect works symmetrically for liberal and conservative decisions. The graphs are the byproduct of logit equations that fit the constraints from Tables 2 and 4. Each allows an interaction effect between gender and opinion. While one plausible outcome would be that while females are more influential when voting liberal, there would be little or no gender difference in influence when voting conservative. In the constrained equations that produce the simulations, male judges are only marginally less influential on conservative votes than females are on liberal votes. In other words, the data suggest the virtual absence of a gender-opinion interaction. This is not obvious from visualizing the graph simulations because conservative votes dominate liberal votes in frequency.

Three Panel Sequence			Two-Panel Sequence Including One with a Liberal-Voting Female Co-Panelist		
Sequence	Panel Type	Adjusted Frequency of Liberal Vote			
Panel 1	All Male	-14.1			
Panel 2	With Liberal Female Panelists	+38.1			
Panel 3	All Male	-6.7			
Panel 3 minus Panel 1		+7.4 (8.2) *p*=.38			