# Appointment Rules and Gender Diversity on High Courts

Nancy Arrington* Leeann Bass † Adam Glynn‡ Jeffrey K. Staton§
Brian Delgado ¶ Staffan I. Lindberg‖

June 30, 2019

**Abstract**

Do the processes states use to select judges for peak courts influence the gender diversity of those courts? Scholars argue that appointment rules are important determinants of the diversity of courts; however, they disagree about the effects of particular design choices. Notably scholars disagree about whether diversity is promoted best by concentrating appointment power in a single individual or by diffusing appointment authority across many individuals. We clarify the mechanisms implied by existing theoretical accounts, derive testable implications concerning the appointment of the first woman to a historically all-male court, and then develop a matched-pair research design within a Rosenbaum randomization inference approach to observational studies. We carry out our analysis on a global sample of high courts beginning in 1970. The evidence is not consistent with the claims that either increasing or decreasing the number of individuals involved in an appointments process promotes gender diversity. Instead we find considerable evidence that a change of any sort in the rules governing appointments promotes diversity, especially changes in the rules regarding the individuals empowered to nominate judges. These results reflect claims that point to institutional disruptions as critical drivers of gender diversity on important political posts.

*Assistant Professor, Department of Political Science, Cal Poly State University San Luis Obispo

†Ph.D. Candidate, Department of Political Science, Emory University

‡Associate Professor, Department of Political Science, Emory University

§Professor, Department of Political Science, Emory University

¶Department of Political Science, Emory University

‖Professor, Department of Political Science, University of Gothenburg

The world's peak courts, by which we refer to high ordinary courts or constitutional courts, are increasingly staffed by women (Hoekstra 2010, Hoekstra, Kittilson and Bond 2014, Turquet 2011). Within the last decade, gender parity or near parity has been reached on the national high courts of Angola, Australia, Canada, Ecuador, Rwanda, Serbia, and Slovenia. Women are increasingly serving as presidents of prestigious courts known internationally for their innovative jurisprudence, including the Supreme Courts of Canada and Israel as well as the Constitutional Chamber of the Costa Rican Supreme Court. This pattern is notable for a number of reasons. The presence of more women on peak courts may influence the law, and by implication, core matters of public policy, either because women understand the law in particular contexts or evaluate facts differently than men (e.g. Boyd, Epstein and Martin 2010, Glynn and Sen 2015, Collins, Manning and Carp 2010) or becasue male judges behave differently when they share the bench with women (Boyd, Epstein and Martin 2010, Farhang and Wawro 2004). It is also possible that more diverse courts promote the legitimacy of the justice system (e.g. O'Connor and Azzarelli 2011, Kenney 2013), and increased gender diversity on important courts may be conceived of simply as an unalloyed normative good (e.g. Malleson 2003). Despite the trend toward diversification, as Figure 1 suggests, most states are very far from gender parity.

Many of the potential causes of the underrepresentation of women on peak courts involve deeply complicated sociocultural processes, including limited availability pools that follow from gendered appointments to lower courts (Anleu and Mack 2009, Epstein, Knight and Martin 2003), elite networks that fail to identify qualified candidates (Linehan 2001), and culturally structured role perceptions that make a prestigious judicial career simply easier for men to pursue (Feenan 2008, Kim 2009). Yet some potential causes are far simpler. Just as scholars interested in descriptive representation among members of parliament have focused on the rules for selecting candidates, judicial scholars have focused on appointment processes. This paper addresses whether particular institutional rules for appointment can help diversify a peak court.

Arguments and findings about the effect of particular appointment institutions are varied. Studies have suggested that executive appointment mechanisms promote greater diversity on courts by promoting clear accountability processes, which facilitates credit claiming for diverse selections (Bratton and Spill 2002, Williams and Thames 2008, Valdini and Shortell 2016). This argument suggests that appointment mechanisms in which multiple actors play a role are subject to coor-

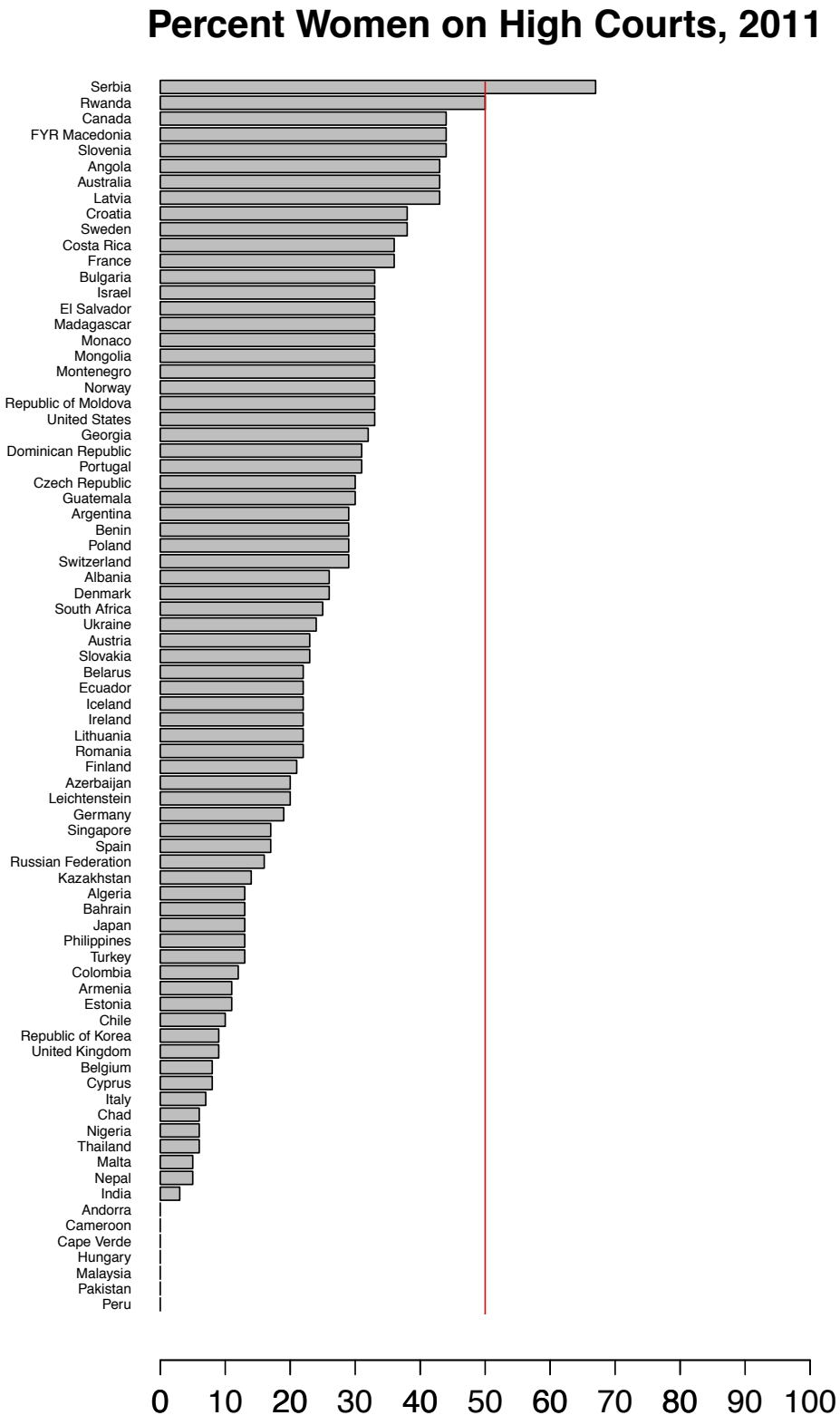


Figure 1: *Gender Diversity on Hight Courts. Data from Turquet et al 2012.*

dination failures as one actor can blame another for the absence of diversify. Thus the key is concentrating appointment power so that whatever political pressure there is for diversification can be concentrated on a single individual.

An alternative argument is that appointment mechanisms that assign multiple individuals to the appointment process are more effective for promoting diversity than are rules that concentrate appointment power in a single office or individual (Gill 2012). This argument deemphasizes coordination failure and instead raises the possibility that appointment rules impact the pools of candidates who are deemed qualified. Adding more actors to the process should increase the chances of diversification by expanding pools of qualified applicants.[1] Further, some scholars have questioned whether appointment institutions matter under all conditions or even at all. In the U.S. case, Bratton and Spill (2002) find that executive appointment to state supreme court positions only promotes greater diversity when courts are all-male. The difference they observe between executive appointment schemes and appointment via elections is insignificant once at least a single woman is on the bench. Hoekstra, Kittilson and Bond (2014) suggest that the diversity of legislative bodies, rather than the precise rules used for selecting judges, are the key to ensuring judicial diversity.[2]

Taken as a whole this literature has developed persuasive theoretical arguments linking appointment processes to diversity outcomes, yet the empirical record is mixed. Even if we wished to suggest an institutional form that promotes more gender diverse courts, the empirical findings have yet to consistently favor one theoretical argument over another; and, in light of the persuasiveness

[1]Still other scholars point to quotas or goals (A summary is found in Hoekstra 2010). In her case for gender quotas for the Supreme Court of the United Kingdom, Malleson (2014) claims "The main argument in favour of quotas is that, unlike all other methods, they are guaranteed to work." Yet, Hoekstra (2010) reminds us that an early attempt to ensure gender diversity on the Supreme Court of Ecuador via a quota did not increase the proportion of female judges. Gender quota rules it would seem, like many rules, can simply be ignored in some cases.

[2]Specifically, they argue that legislative gender quotas affect high court diversity by creating diverse legislatures, which help push for greater diversity in justice institutions. The effect of quotas, mediated by legislative gender diversity, is independent of the rules that are used for appointments.

of the arguments, we are somewhat at a loss as to which institutional form to recommend. In light of the normative issues at stake, we believe that it is essential to seek credible answers to questions about the potential effects of appointment institutions on a peak court's gender diversity; however, we also believe that existing arguments can be pushed further and integrated, the result of which is to reveal new empirical implications and to highlight important concerns of research design.

Our study contributes directly to research on judicial appointment institutions and gender diversity; however, in so far as the challenges we confront manifest in many areas of institutional research, our theoretical approach and research design are broadly applicable. We highlight three contributions. First, we further develop the two central competing theoretical arguments linking institutional rules for appointments and diverse benches. A key feature of institutional research is that the rules that structure political choice are typically linked to underlying political context, e.g., individual preferences, group structure, culture, etc. (e.g., Neto and Cox 1997, Riker 1982). Drawing on the work of Bratton and Spill and Hoekstra, Kittlson and Bond, we develop institutional models of appointments in which a state's underlying pressure for institutional diversification influences institutional effects. We will show theoretically that once underlying pressure for reform is incorporated into either of the main arguments scholars have developed, institutional effects are strongest at moderate levels of pressure. We also show theoretically that the effects of an institutional change depend on the very institution that has been changed. In this sense, institutional effects are path dependent. As we discuss below, these points not only increase the set of empirical observations to observe, they have critical implications for research design by structuring the very comparisons we should be making.

Second, we develop an empirical design that permits credible causal claims regarding the empirical implications of existing arguments, a design that can be used by scholars working on related problems in fields other than gender and judicial institutions. Leading studies in the field rely on cross-sectional designs and linear or logistic regression (e.g., Williams and Thames 2008, Bratton and Spill 2002). Although these designs are potentially valid, as Hoekstra (2010) and Hoekstra, Kittilson and Bond (2014) note, they rely on a number of assumptions that may not hold in practice. The most obvious concern is whether the designs control for measurable confounders. Even if they do, without balanced samples, inference is heavily dependent on modeling choices. A deeper yet related concern is that differing adoption times for appointment systems greatly complicate at-

tempts to control for confounding. Consider the United States, which adopted its system in 1789: any control variable measured after 1789 could potentially induce post-treatment bias with respect to the United States. Generalizing this intuition to a global sample, any control variable measured recently could induce post-treatment bias with respect to many countries. In addition, considerable differences in the sizes of peak courts complicates the use of an outcome variable like the proportion of judges on a court who are women. Most obviously, a change from two to three female judges means something considerably different on a five member court than it does on a court with twenty judges or more. And even if we adopt the appealing normative position that courts should reflect the societies in which they are embedded, without precise estimates of availability pools across all states and times, it is not clear exactly what proportion of a court would reflect the optimal level of descriptive representation.[3]

Our design addresses these concerns in a number of ways. We leverage changes in constitutional processes for appointments to a state's highest ordinary court or constitutional court if one exists. To ensure that we are making sensible comparisons, we then exact match each state that had a constitutional change to a state that did not. In exactly matching, we ensure that states without changes had the same institutional framework as the state that changed, in the year of the change. We then further match the sample with respect to a number of pre-treatment confounders suggested by the literature. To ensure that we can validly interpret changes in descriptive representation on peak courts, we focus on states whose courts were entirely staffed by men at the time of the institutional change. We then measure the time between the constitutional change and the year in which the first woman was appointed. This design greatly clarifies the implicit comparisons that are made in the typical regression approach. As we discuss below, it is also clearly linked to our theoretical models. That said, what we gain in the clarity of research design we lose in sample size. To address this issue we adopt a Rosenbaum (2002) randomization inference-based approach, which allows for non-parametric inference in small samples. The approach has been used in political

[3]The exception here is when the availability pool is homogenous – such as before women were legally allowed to attend law school or hold political posts. In these cases, the expected level of descriptive representation of gender on courts would be no descriptive representation.

science to evaluate the effects of electoral rules on opposition repression (Glynn and Ichino 2015) as well as the effect of ballot order on candidate success (Ho and Imai 2008).

Third, our findings suggest that appointment structures cause delays in the appointment of women to historically homogenous courts; however, not in the ways suggested by either of the dominant models of appointments evident in the literature. Our analysis does suggest some evidence in favor of the key empirical implication associated with each theoretical argument: an increase in the number of actors involved in the selection process leads to slightly faster diversification but so does a decrease in the number of actors. And the results are stronger in contexts where we believe there would have been moderate pressure for reform. While suggestive in our small sample, neither result is strongly statistically significant. More importantly, as we discuss below, this pair of findings is simply inconsistent with both dominant theoretical models. What we find instead is evidence consistent with an argument that highlights institutional disruption as key for diversification (Darcy and Choike 1986, Hughes 2007). We find very strong support that any change in the institutional processes by which peak court judges are selected causes all-male courts to diversity. This effect is concentrated on the individuals and groups in a state who are responsible for proposing candidaes. Although these findings do not rule out the possibility that coordination failures or diffuse accountability rules do in fact matter, they strongly suggest that, at least in the context of all-male courts, if you want to increase gender diversity, you would do well by shaking up your selection process in whatever way seems most sensible.

The remainder of the paper is organized as follows. We first develop a theoretical framework that formalizes existing arguments about the effects of appointment institutions found in the literature. We then describe our research design, present our data, and summarize our results. Finally, we consider what these results suggest about cross-national research on appointment processes in the context of judicial diversification.

## Linking Appointment Rules to Diversity Outcomes

Our goal in this section is to further develop two existing arguments that link judicial selection institutions to diversity on the bench by (1) broadening the search for qualified candidates or (2) reducing the possibility of coordination failure. We describe classes of appointment processes for

peak courts and place those processes in the context of existing theoretical arguments. We then present two simple models of appointments that attempt to capture the core logics of popular theoretical perspectives on appointment institutions, while simultaneously linking those logics to underlying social demand for greater diversity on the bench.

### Key Elements of the Appointments Process in Prior Work

Institutional scholarship on gender and peak courts has focused on three aspects of the process – the identification of qualified candidates, the nomination of particular individuals and the decision to assign an individual to a particular seat. Over the past five years, our research team, associated with the Varieties of Democracy Project (V-Dem), has been collecting information on appointment and removal institutions for peak courts. Working in cooperation with the Comparative Constitutions Project (Elkins, Ginsburg and Melton 2012), we are developing descriptions of the appointment and removal institutions for peak courts in all states (and many former colonies), from 1900 to the present. Constitutionally defined procedures largely, though not always, focus on two stages. Formally, the identification of candidates as well the development of particular proposals takes place via what we will refer to as the "nominations stage" (For a careful discussion of the aspects of the appointment process, see Gill 2012), where a legally empowered actor proposes a name for a seat. We will refer to each institutional actor empowered to name a candidate for a position on the court as a "nominator." This is often a president, in which case it refers to a particular individual. It can be a chamber of a legislature or a commission, which are clearly groups of people. By nominators, we refer to both types of actors. The second stage, what we will refer to as the "appointments stage" involves a legally empowered actor assigning a proposed name to a seat. We refer to these actors as "selectors," again allowing the term to describe individuals as well as institutionally identified groups. We refer to the sum of nominators and selectors involved in a particular process of appointing a judge simply as "actors."

In practice, the nomination and appointment stages can be combined in a variety of ways. Figure 2 describes four distinct constitutional processes for appointing constitutional court judges. The Peruvian process, depicted at the bottom-left of the figure, is clearly the most simple. The Constitution of Peru creates a process in which nomination and appointment are assigned to the same institutional actor, the Peruvian Republic's Congress. The process for selecting Federal

Constitutional Court judges in Germany is marginally more complex, in that the upper and lower chapters of the German parliament are empowered to select one half of the judges who sit on the Federal Constitutional Court. As in Peru, within the chambers, nominating and appointing authority are combined. Both of these processes contrast with that used to appoint members of the Mexican Supreme Court and the Constitutional Court of Serbia in that the nominations and appointments stages are assigned to distinct actors. Mexico's process involves presidential nomination followed by appointment by the Senate for each of the Court's eleven justices. The Serbian process both separates the nominations and appointments stages and assign powers to multiple actors. Specifically, five judges are appointed via one one of the following three processes: (1) the National Assembly nominates and the President appoints, (2) the President nominates and the National Assembly appoints, and (3) the High Court and Prosecutor's Council nominates and the Supreme Court of Cassation appoints.

To clarify the meaning of our terms, Peru has only one nominator and one selector; however, we will say that there is only one institutional actor involved in the entire process, since the stages are combined. Germany has two nominators and two selectors but also only has two institutional actors in the entire process. Mexico has a single nominator and a single selector, but since they are different institutional actors, in contrast to Peru, Mexico has two institutional actors in the appointments process. Finally, Serbia has four institutional actors involved in appointments, as well as three nominators and three selectors.

Scholars who value group appointment processes highlight the positive effects on the identification of qualified women that follows from tapping into a more varied and dense network of contacts (Gill 2012). The Serbian process reflects well the kind of procedure that we might envision under this approach. Not only does final appointment power reside with multiple institutional actors (and many individual people), so does the nomination stage as well. Scholars who see the potential for coordination failure among larger groups of decision-makers envision rules that concentrate power in a small set of actors. Under this approach Mexico's presidential system reflects a natural process, yet the German approach is similar in so far as blame can be placed on two institutional actors. In principle, the Peruvian system is the most concentrated mechanism for appointment and thus

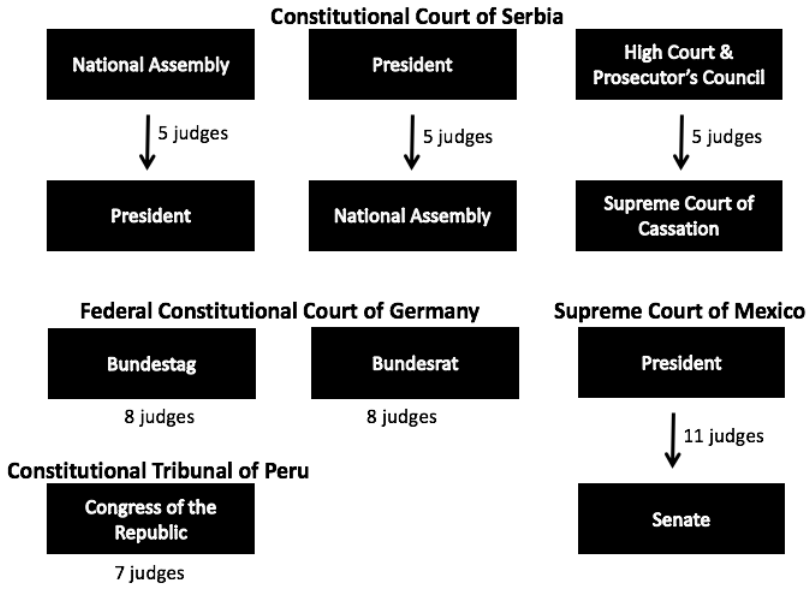


Figure 2: *Constitutional processes for appointing constitutional court judges in Serbia, Germany, Mexico and Peru. To allow a comparison to Figure 2, the processes described in the figure reflect those in practice in 2011.*

the most likely to produce diversity if coordination is the problem.[4] In practice, women constitute over a majority of the members of the Constitutional Court of Serbia. The German Federal Constitutional Court is roughly one-third female as is twenty percent of the Supreme Court of Mexico. The Constitutional Tribunal of Peru has one woman. This very small sample indicates support of the notion that more actors should promote diversification. Whether this pattern should hold and under what conditions is the subject to which we now turn.

## Modeling appointments for gender diversity

The empirical implications of existing arguments seem clear, but we believe that there are important reasons to formalize their logics. As we will show, these logics suggest a larger, more nuanced set of implications than have been heretofore recognized, and which raise important challenges for standard approaches to inference. Our particular aim then is to develop the logic of existing argu-

[4]Note: There is a complication here in so far as the fragmentation of a legislature (and obvious concern in Peru) might render the Peruvian process less concentrated.

ments when underlying sociopolitical features of a state are incorporated, which for simplification we will refer to as "pressure" for a diverse bench. Of course, this might be done in a variety of ways, via a variety of classes of models.

The literature on Supreme Court appointments in the United States conceives of the process as one involving inter-branch bargaining (e.g. Moraski and Shipan 1999). Appointment processes generally can be understood fruitfully as a bargaining process. We will depart from that framework here, because existing arguments about gender diversity do not envision a bargaining setting. Rather the focus is either on the ways that appointment rules change the possibility for holding officials accountable or on the ways that rules incentivize more creative searchers for qualified judges. Scholars who highlight the advantages of group selection seem to have in mind a process that taps into independent and diverse information networks. The basic idea is that by involving more entities in the search and appointment process, the information network expands, which increases the chances of a successful search for diversity. In contrast, scholars who see group processes as potentially problematic seem to have in mind a coordination problem that can emerge when choices are interdependent. To summarize then, rather than a bargaining problem, scholars of gender diversity seem to have in mind something closer to a public goods provision problem, where the good being supplied is greater gender diversity on a peak courts.

Given our goal to develop existing arguments, we will conceive of appointment rules for increasing gender diversity as influencing a kind of public goods provision problem. The key difference between existing arguments is whether choices are considered as independent or interdependent. Scholars who argue for a large number of actors to be included in the appointment process seem to have in mind a process in which decisions are made independently. This independence implies rather directly that an increase in appointers should make it increasingly likely to add diversity. In contrast, scholars who promote single appointers envision choices as interdependent. This interdependence raises the possibility of coordination failures as the number of appointers increases. Our approach will be to develop models that follow from both assumptions. As we hope to make clear, the clearest interpretation of the models we consider is that they illuminate challenges in the development of candidates for consideration. This is to say that existing arguments in the literature, which highlight coordination failures and clear lines of accountability are best thought of as models of the process by which shortlists of judges are developed. Absent a clear account of

bargaining, implicit in these arguments is that there is a positive relationship between the number of women on a shortlist and the probability that a historically homogenous court will diversity. We adopt this assumption as well.

## Two Models

We assume that an appointment processes can be described by an integer, $n \geq 1$, which represents the number of individuals empowered to appointment a judge to a court. Each individual chooses from the set $G = \{M, F\}$. We thus collapse the nominations and appointments stages and place our emphasis on the number of institutional actors capable of choosing a judge. We view this choice as sensible in light of the very large number of bargaining models that might apply to our setting, in which selectors can be highly advantaged relative to nominators or vice versa. A slate of appointed judges is characterized by a list $l = \{g_1, g_2, \ldots, g_n\}$ of judge genders $g_i \in G$ for selectors $1, \ldots, n$. We will say first that an appointment process is considered a success if at least one member of $l$ is $F$. Finally, we will assume that there is pressure in society for the appointment of a woman, which we will denote $v \in (0, 1)$, and where values of $v$ closer to 1 indicate stronger pressure for a female appointment.

In order to model the dynamic envisioned by models that privilege a larger number of appointers, we consider a process in which appointer choices are independent of each other. To model the dynamic envisioned by models that privilege a smaller number of appointers, we consider a process in which these choices are interdependent.

### Independent Appointments

Consider a simple probability model of choice. Suppose that the choices of the selectors are linked to underlying pressure for diversity. Specifically, let the probability selector $i$ sets $g_i = F$ be $v$, so that selectors are increasingly likely to choose a female judge as pressure for diversity increases. The number of female judges appointed via this process is thus a random variable with distribution $B(n, v)$, so that the probability of $k$ women being selected in an appointment process is given by

$$\binom{n}{k} v^k (1 - v)^{n-k}.$$

Given our definition of success, we are interested in the probability of at least one woman being appointed in a search. This probability is

$$1 - \binom{n}{0} v^0 (1-v)^n = 1 - (1-v)^n.$$

Importantly (and transparently) the probability of success increases monotonically in $n$. This very simple model clearly captures the basic logic of the argument that adding additional selectors to the process increases the probability of a successful search. Indeed, as the number of selectors increases without bound, the probability of success approaches 1. But the model does more. It clarifies one way in which rules might interact with underlying pressure for diversity. At very high demand, where every selector is likely to identify and appoint a qualified woman, adding selectors has a very small impact at the margin. Substantively, there is such high pressure for diversity in this setting that even a single selector would have ample incentives, information and opportunities to ensure a successful search. For a similar reason, where demand is very low, adding an additional selector, while certainly increasing the probability of a successful search, will have a relatively low impact. There simply are too many ways to fail to identify a qualified female candidate when pressure is too low. It is at moderate levels of pressure where institutional form should make a meaningful difference. Table 1 shows a few examples for values of demand reflecting low, middle and high levels, which summarize the following proposition.

| Pressure ($v$) | One Selector | Three Selectors | Four Selectors | $\max(\triangle\ Pr(Success))$ |
|---|---|---|---|---|
| .02 | .02 | .04 | .07 | .05 |
| .50 | .50 | .875 | .94 | .44 |
| .98 | .98 | $\approx 1.0$ | $\approx 1.0$ | .02 |

Table 1: *Probability of a success with increases in the number of selectors, given sociopolitical demand ($v$) for diversity. Column 1 shows the level of sociopolitical pressure for an increase in diversity. Columns 2 through 4 show probabilities of success for selectors $n = 1, \ldots, 4$. Column 5 shows the difference in probability of success for four selectors vs. one.*

**Proposition 1** *The probability that a court has at least one woman judge is increasing in the number of appointers, but the effect is strongest at moderate values of demand for diversity, relative to low or high values.*

**Interdependent Appointments**

We now consider a simple game theoretic model of the appointment process, which we believe reflects the basic concern of scholars who have promoted the executive appointment mechanism. The intuition we wish to capture is that pressure for adding gender diversity will be reduced as we increase the number of appointers, permitting each appointer to rely on the others. The simplest form of it we can identify is a textbook public goods provision game, which is elaborated by Osborne (2003) among others. We model this process as an $n \geq 2$ player public goods provision game, in which each player benefits identically from a successful search. Let the utility value of success be $v$, so that the value increases in underlying sociopolitical pressure for diversity.[5] We also assume that each appointer confronts a personal (but identical) opportunity cost, $c \in (0, 1)$, for selecting the female candidate. This cost might be very small, as when the female candidate under consideration is strongly preferred to the male alternative, for reasons unrelated to gender. The cost might be relatively large when the opposite is true. A pure Nash equilibrium in this game is a list $l$ of $n$ judge choices. For sufficiently high demand for diversity, i.e., $v \geq c$, there are $n$ lists that are equilibria in this game. Each list involves a single appointer choosing $g_i = F$, where all other appointers choose $M$. In contrast, if $v < c$ there is only one equilibrium in which all $g_i = M$. In both cases, the institutional form has no effect on outcomes – either the diversity search will be successful or it will not be successful, but this all depends on society's demand for diversity, in a sense reflecting the essential logic of Hoekstra, Kittilson and Bond (2014).

This version of the model, of course, fails to capture the underlying intuition of concentrating appointment power on a single person. Scholars must have a different dynamic in mind. Consider the symmetric mixed strategy Nash equilibrium of this game, where each appointer appoints a

[5]Note, all appointers obtain $v$ if there is a success. An alternative model might consider scenarios in which appointers are only rewarded for personally increasing diversity. If this is the model though, it is hard to see the value of considering a group choice.

female candidate with probability $p_i$. A mixed strategy Nash equilibrium in this game is a list of $n$ probability distributions over $G$, such that that $p_i = p_j$ for all players $i$ and $j$.[6]

For this to be an equilibrium with all $p_i \in (0, 1)$, we first require $v \geq c$, as before. If this does not hold, all players will set $p = 0$. Assuming that this condition holds, it must be that each appointer is indifferent between appointing a female and a male judge. If this condition does not hold, given the symmetry of strategies, then either all appointers choose a male or a single appointer will choose a woman, while all others choose a male; and, thus, the institutions we use will be irrelevant.

Recognizing that the utility value of appointing a woman is $v - c$ for each appointer, for equilibrium we require that this value is equal to the expected value for each appointer a choosing a male judge. This expected value is the utility of at least one woman being chosen (i.e., $v$) multiplied by the probability that at least one other appointer chooses a women. For this kind of equilibrium, we require the following to hold for every appointer:[7]

$$v - c = v \cdot (1 - (1 - p)^{n-1}) \tag{1}$$

$$\frac{c}{v} = (1 - p)^{n-1}. \tag{2}$$

Before proceeding it is useful to note that right hand side of (2) reflects the probability that no other selector will pick a female judge. The key point is that, in equilibrium, this probability is independent of the number of appointers ($n$). It simply reflects the ratio of opportunity costs to the social pressure for diversification. When $c$ is very large relative to $v$, clearly no female judge will be appointed by the appointer in question or anyone else. In contrast when social pressure for diversification is extremely strong relative to opportunity costs, there will be at least one female appointment no matter what any one appointer chooses. Importantly, in a symmetric

[6]In this case, i.e., where strategies are symmetric, we are assuming that every appointer picks an identical distribution, i.e., each appointer chooses the same probability of appointment a female judge.

[7]Since every appointer chooses the same $p$ in equilibrium we suppress the index.

mixed strategy Nash equilibrium, each selector should set the probability of naming a female judge as follows:

$$p^{\star} = 1 - (\frac{c}{v})^{\frac{1}{n-1}}. \tag{3}$$

There are three key points to consider. First, note that as $n$ increases without bound, the probability with which each appointer selects a woman approaches 0. Adding appointers incentivizes each appointer to more significantly rely on the other appointers in order to to solve the group goal. Second, the probability that no female judge is appointed is just the probability that the appointer in question selects a male judge multiplied by the probability that no other appointer selects a female judge or

$$(1 - p^{\star})(\frac{c}{v}). \tag{4}$$

Given that the probability that all the other $n-1$ appointers fail to select a woman is independent of $n$, it follows immediately that the probability that no female judge is selected rises as the number of appointers $n$ increases. Not only does each individual judge set a lower and lower probability of appointing a woman as the group size increases, the probability that the group as a whole adds diversity also decreases as group size increases. As the essential logic of the scholars who value single appointer processes suggests, coordination failures get worse as the number of appointers increase.

Third, and we think most importantly, the negative effect of adding appointers on the probability of a successful appointment process depends on demand for diversity. In contexts of very strong pressure for diversification, say where $v$ increases without bound, note that the second term in (4) approaches zero. Thus, independently of the effect of $n$ on the probability of one appointer's choice, the total probability of an unsuccessful search will approach 0 – courts will diversify. Similarly, we already know that if the opposite relationship holds, and $v$ is smaller than $c$, then the process will simply not result in added judicial diversity no matter the number of appointers.

**Proposition 2** *For low pressure for diversity, adding appointers to an appointment process has no effect on the probability of having at least one woman on a court. For sufficiently high levels of pressure, adding appointers (weakly) decreases the probability of a female appointment. Assuming*

*that pressure is sufficient to increase diversity at all, the effect of adding appointers is strongest at moderate levels of demand, relative to either low or high demand.*

### Summary of findings

Whether you conceive of the process as involving essentially independent or interdependent choices, the effect of adding an additional actor to the process depends on social pressure for an increase in diversity. Figure 3 summarizes these results graphically. The top panel plots the change in probability of at least one female judge being appointed associated with changing from a single appointer process to a process with two appointers,[8] as a function of pressure for diversity ($v$). The bottom panel shows the same results for the model in which choices are interdependent. The graphs are different in a few obvious ways. First, the effects described in the top panel are all non-negative whereas they are all non-positive in the bottom panel. And of course the shapes are clearly different. Notably, the bottom panel illustrates that at very low levels of pressure for diversity, the number of appointers does not affect the diversity of the appointments. The discontinuity in the bottom panel reflects difference in the probability of a successful search at relatively low pressure, where a single appointer would select a female judge but a pair of appointers would not. Most importantly, both panels indicate that the largest effects (positive or negative) are at moderate levels of pressure.

A final point of comparison between the models in described by Figure 4, which shows changes in the probability of appointing a woman associated with a one appointer increase in the process, for all institutional processes prior to the change. The left panel shows these effects for the independent choice model; the right panel shows the same effects for the inter-dependent choices panel. The key point is that while changes in the effects of an increase in each model are monotonic, they are decreasing. The biggest effects are for changes in very small appointment groups. Considering the left panel, the figure suggests that the change in the probability of success is nearly 0.25 when a one appointer process is changed to a two appointer process. This effect drops all the way to 0.10

[8]Specifically, we plot $Pr(k \geq 1|n = 2) - Pr(k \geq 1|n = 1)$, where $k$ is the number of women appointed in a process and $n$ is the number of apponters.

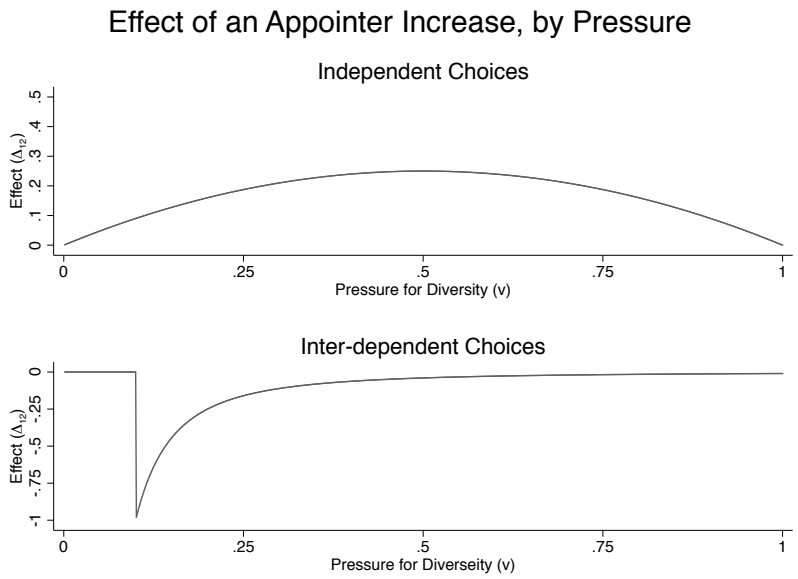


Figure 3: *Shows the change in the probability of at least one female judge being appointed associated with a single appointer increase, across the complete range of pressure for diversity (v). The top panel shows results for the independent appointments model. The bottom panel shows results for the interdependent appointments model. Both cases show the results for a change from one appointer to two.*

when we consider the effect of changing from a two appointer process to a three appointer process. This heterogeneity will guide our empirical approach.

## Research Design

Our research design combines pair matching and Fisherian randomization inference for matched pairs. The procedure involves the following steps. We first identify states that have had changes in their constitutional articles which define the process for appointing judges to a peak court. We refer to these states as *treatment states.* We exactly match each treatment state to potential control states using pre-treatment institutional information on the number of nominators and selectors in the appointments process and the year of constitutional change.[9]

Third, among the states that are exactly matched on institutional structure, we use propensity score matching to identify pairs of treatment and control states on the basis of likely pre-treatment

[9]We also consider using the total number of nominators or selectors.

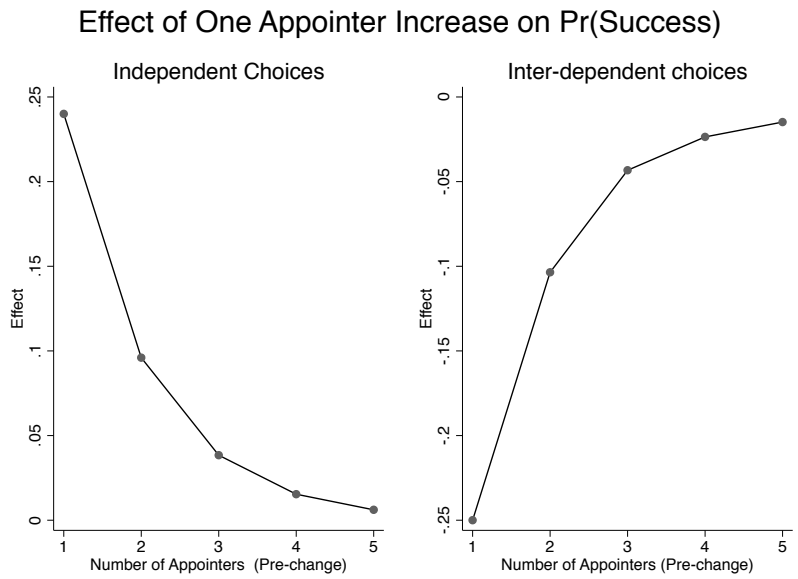


Figure 4: *Shows the change in the probability of at least one female judge being appointed associated with a single appointer increase, for different numbers of appointers prior to the change. Specifically, then we consider effects associated with one appointer increases using 1, . . . , 5 as the baseline number of appointers. The left panel shows results for the independent appointments model. The right panel shows results for the interdependent appointments model.*

confounders. This process results in the matching of each treatment state to a plausible control states that is as similar as possible on pretreatment confounders and exactly matched with respect to the institutional structure for selecting peak courts judges in the year that its treatment state experienced a constitutional change. For each pair, we calculate the difference in time to the appointment of the first female judge between the treatment and control state, following the year of the constitutional change for the treatment state. We then calculate a signed-rank statistic for these differences. Finally, we compare this statistic to the permutation distribution of signed-rank statistics derived from our sample. We test the sharp null hypothesis that treatment is independent of the potential outcome under the treatment and control condition for every unit in the study, i.e., that a change in the constitutional structure for appointing judges is not a cause of the time to the appointment of the first female judge for any state we consider.

### Outcomes and Treatments

As we just described, the outcome of interest is the time in years from an institutional change the appointment of the first woman to a peak court. Using text provided by the Comparative Constitutions Project (CCP), we have diagramed the appointment processes for all states' peak courts from 1900 to the present. In so far as we are coding each constitutional event over time, we are able to identify when and how the appointment processes for peak court judges changes.[10] Our study focuses on changes made between 1970 and 2000. Figure 5 illustrates what we understand to be a treatment state. The Constitution of Nicaragua was reformed in 1987 as a part of the institutionalization of the Sandanista regime. A part of the reform included a change in the appointment process for the Supreme Court. Like modern day Peru, Supreme Court justices were nominated and appointed by the National Assembly from 1962 through 1986. The 1987 reform assigned nomination power to the President leaving appointment power in the National Assembly. This change reflects an increase from one actor in the appointment process to two actors. We thus can treat Nicaragua in 1987 as a treatment observation, where treatment is understood as an increase in the total number of actors in the process.

We focus on the total number of actors for two reasons. This represents the closest connection to the theoretical model. It also avoids a series of difficult choices over whether and if so how to weight the roles that nominators and selectors play in this process. In some settings, the nominators will be highly advantaged in a bargaining process but in others it will be the selectors. In so far as disadvantaged bargaining partners anticipate the reactions of the advantaged partners, it may be nominators propose to selectors exactly whom selectors want. Absent an account to guide us in identifying exactly the actors who matter in every process, we are conservative and simply count the total number of nominators and selectors in the process. That said, we will consider the robustness of our findings to alternative conceptualizations of the treatment.

[10]The limited exception is for a small set of countries for which CCP did not have a constitution and where our research team was unable to translate it. For constitutions not yet translated, we have only been able to identify changes in the constitutions written in languages our coding team can read: English, Spanish, French, Portuguese, Dutch, Polish, Russian, Romanian, German, Finnish and Norwegian.

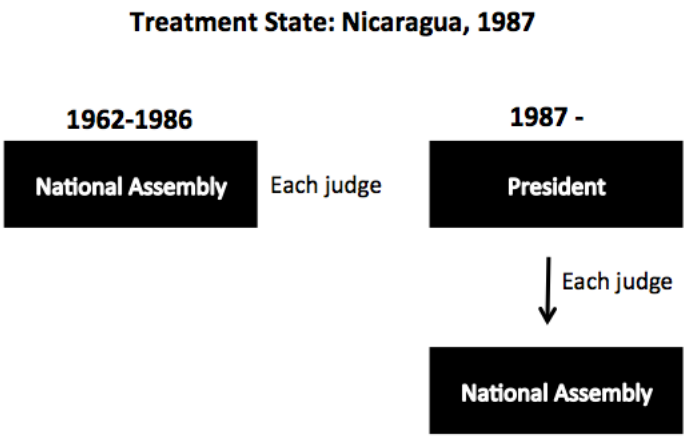


Figure 5: *Illustrates nominations and appointments process for the Supreme Court justices in Nicaragua. The process involved a single actor, the National Assembly, from 1962 to 1982. In 1987 the Constitution was reformed and assigned the nomination power to the President, leaving the appointment power in the National Assembly.*

## Matching

The models that guide us indicate that the effects of a change in institutional form depend on not only the pre-treatment level of political pressure for gender diversification but on the existing institutional framework prior to the change. To ensure that we can validly interpret a change in the number of women on a court as successful with respect to gender diversification, it is useful to focus on courts that had no women at the time of the institutional change. We thus match every treatment state to a state that had the same number of nominators and appointers in the year that treatment state's institutions changed. We also ensure that all states in the study had yet to have a woman appointed to their peak court at the time of treatment. We conduct this exact matching in the CEM package in R (Iacus, King and Porro 2009).

Within the set of exactly matched states, we then use propensity score matching in the package Optmatch to generate matched pairs (Hansen and Klopfer 2006). Because the effect of institutional changes on court gender diversity is conditioned by the existing social pressure for diversity, we include several potential confounding variables in the analysis. First and foremost, we include the

percent women in the lower house of a country's legislature or parliament.[11] As Hoekstra, Kittilson and Bond (2014) find, there is a relationship between the level of gender diversity in the legislative branch and subsequent gender diversity on high courts. It is possible that both the level of gender diversity in legislative office and on high courts is shaped by outside social forces, but there is some evidence that the presence of women in legislative office might lead to increased presence of women on courts (Hoekstra, Kittilson and Bond 2014): the presence of women in legislative office can serve as a signal to judicial selectors that women are already involved in the political process and that citizens value gender diversity in office.

We also seek to condition on the amount of time since universal suffrage was granted for each country (also from Paxton, Green and Hughes (2008)). We expect that the longer women have had access to political power through voting, the pressure for increased access to institutional positions of power for women will also be greater. To capture a broader set of features that are associated with social pressure to diversify, we also include a Women's Political Empowerment Index from the Varieties of Democracy project. This index was created through point estimates of a Bayesian factor analysis model of several variables: women's participation in civil society participation, percent female journalists, freedom of domestic movement for women, freedom of discussion for women, freedom from forced labor for women, property rights for women, access to justice for women, and power distributed by gender (Coppedge et al. 2015). All of these measures should be associated with an underlying social pressure for or preference for gender diversity on peak courts.

In addition to the Female Empowerment Index, we include several additional indices and measures from the Varieties of Democracy Project. The Egalitarian Democracy Index measures the extent to which the ideal of egalitarian democracy– that is, the extent to which there is an equal distribution of political power across social groups such as class, sex, religion, and ethnicity–is

[11]Data are from the InterParliamentary Union, but we used versions collected by (Paxton, Green and Hughes 2008) for years 1970-2003 and (*Proportion of seats held by women in national parliaments* N.d.) for years 2003-2010. Missing data between two observed data points were filled in with liner interpolation. In addition, if there were no women in legislative office, preceding missing values were filled in, likewise, as zeros.

achieved in a given country year. Similarly, the Participatory Democracy Index measures the extent to which all citizens can be active in the political process.

We also include a series of variables from the Varieties of Democracy Project. Specifically, we use a measure of court packing; a measure of the extent to which women have access to justice; a measure for whether and the extent to which power is distributed by gender; a measure of women's civil society participation; a measure of the presence of female journalists; a measure of the freedom of domestic movement for women; a measure of the freedom of discussion for women; a measure of the extent to which women are free from forced labor; and a measure of the extent of property rights for women. Table 2 lists these variables and summary statistics for the full, unmatched data set.[12]

| **Variable Name** | **Mean** | **St.Dev.** | **Range** |
|---|---|---|---|
| Year | 1991 | 15.55 | 1960, 2015 |
| Number of Nominators | 0.55 | 0.96 | 0, 9 |
| Number of Appointers | 1.26 | 0.79 | 1, 6 |
| Percent Women Lower House | 11 | 9.87 | 0, 63.8 |
| Years since Univ. Suffrage | 41.81 | 23.5 | 0, 122 |
| Court Packing Index | 0 | 1.09 | -4.21, 3.46 |
| Freedom of Discussion for Women | 0.5 | 1.62 | -3.55, 3.87 |
| Property Rights for Women | 0.81 | 1.33 | -3.74, 2.93 |
| Women's Civil Soc. Participation | 0.71 | 1.16 | -3.06, 3.24 |
| Power Distributed by Gender | 0 | 1.12 | -2.98, 4.2 |
| Women Political Empowerment | 0.58 | 0.22 | .09, .96 |
| Egalitarian Democracy | 0.36 | 0.26 | .02, .92 |
| Access to Justice for Women | 0.46 | 1.47 | -3.83, 3.55 |
| Freedom Domestic Mvmt for Women | 0.97 | 1.33 | -4.52, 3.53 |
| Freedom from Forced Labor for Women | 0.81 | 1.08 | -3.94, 2.81 |
| Female Journalists | 28.9 | 13.68 | 1.75, 66.72 |
| Participatory Democracy Index | 0.29 | 0.22 | 0.01, 0.84 |

Table 2: *This table shows summary data for the full, unmatched data set. Treated countries will be matched* exactly *on Year, Number of Nominators, and Number of Selectors. Then within exact match subclass, treated countries will be matched to one control country with propensity score matching on the other variables.*

**Matching Consequenes** Figure 6 helps evaluate the consequences of matching. It shows the standardized differences in means across treated and control countries for the unmatched data and

[12]We used linear interpolation to fill in any missing values in years between observed data points.

for matched samples. The open circles indicate standardized differences in means (control means subtracted from treatment means divided by pooled variance) for the unmatched sample. The black points show the same information for the matched sampled data. As expected, exact matching on the number of nominators and appointers as well as the year produces perfect balance. Although there was good balance in the unmatched sample with respect to the number of appointers, treated states had far fewer nominators on average. Thus, the unmatched sample would not permit valid comparisons. Although balance is not perfect for the remaining variables, it is significantly improved for the most significant potential confounder, the percent women in the lower house of the legislature.

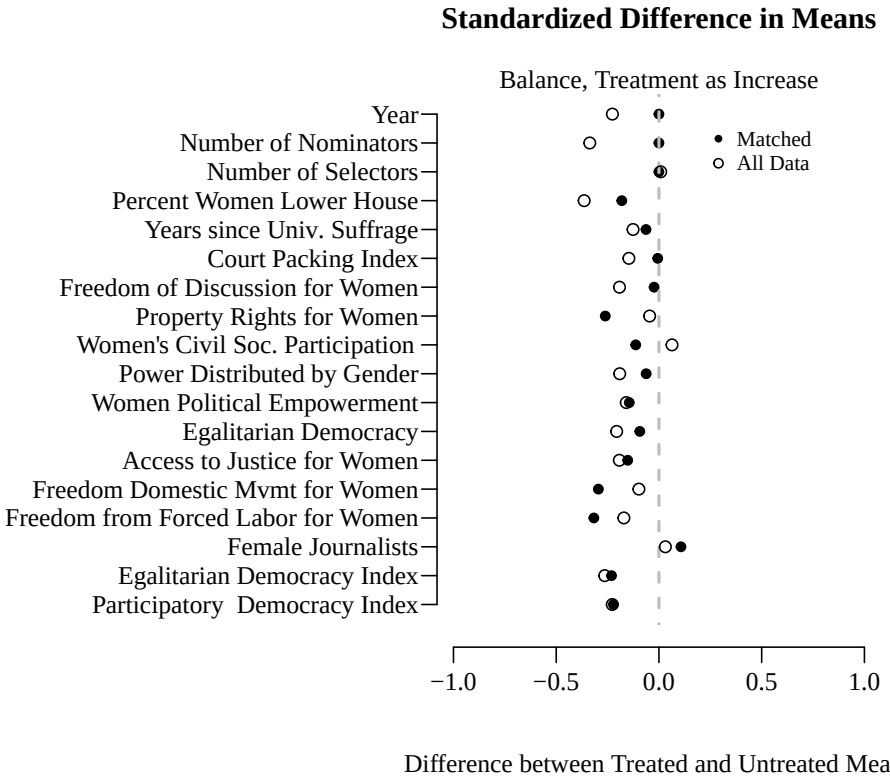


Figure 6: *Displays the standardized difference of means for treated and control units in the original and matched samples.*

Matching ensures that we are making sensible comparisons in our analysis, comparisons that are interpretable in light of existing theoretical logics. Yet, there are costs of ensuring valid comparisons. Specifically, the costs lie in a reduced sample size. Table 3 provides a summary of the consequences

of our matching approach. The first column of the table shows the set of state years in which there was an increase in either the nominators or appointers in a selection process for peak courts. The second column displays the set of control states for each treatment state that satisfied the criteria for inclusion in our study, most importantly the criterion that there had not been a woman appointed to the court in the year of the constitutional change.[13] The third column of the table shows the set of states that had a constitutional reduction in the number of nominators or appointers, as well as the matches for those states. The key point here is that there are at most 16 pairs of states for any analysis.

The matching process also makes clear two features of this kind of analysis that is blurred in cross-national regression analyses. As Table 3 highlights, treated units largely come from the developing world. What we can know about about this process from leveraging changes in constitutional terms will be drawn primarily from developing states, because these are the states whose constitutional terms tend to change. In addition, our comparisons across institutional forms require us to compare states across regions, so that for example, Albania in 1976 is matched to Japan in 1976. Exactly matching on region is infeasible.

After matching we are left with a sample of pairs of states. Recall that the outcome of interest is the time in years from institutional change until the appointment of the first woman to the court of interest. We are interested in whether treatment increases this time; however we are especially interested in the effect for moderate levels of pressure for diversification. Following Hoekstra, Kittilson and Bond (2014) we use the percentage of women in the lower chamber of a state's legislature to measure this pressure. To focus on states with moderate degrees of pressure, where we should see the strongest effects, we subset the data to states that fall within the second and third quartiles of the distribution of percentage women in the lower chamber.

## Analysis

To conduct our analysis, we calculate a signed-rank statistic for each sample. For each pair of states, we know the number of years from treatment until the first woman appointment. We can then rank these differences. We use the statistic as calculated in Glynn and Ichino (2015). In this approach,

[13] And additional criterion is that we have data for the outcomes for each member of each pair.

| **Treated: Increase** | | **Treated: Decrease** | |
|---|---|---|---|
| Treated | Control | Treated | Control |
| Albania 1976 | Japan | Albania 1991 | Paraguay |
| Cent. Afr. Rep. 1994 | Mozambique | Benin 1970 | Liberia |
| Chad 1989 | Suriname | Burundi 1992 | Bahamas |
| Chile 1980 | Tunisia | Nigeria 1978 | Algeria |
| Ecuador 1983 | Guyana | Suriname 1982 | South Korea |
| Kazakhstan 1995 | Croatia | Cambodia 1976 | |
| Lithuania 1992 | Myanmar | Cameroon 1975 | |
| Madagascar 1975 | France | Cape Verde 1999 | |
| Mauritania 1991 | Estonia | Chad 1996 | |
| Niger 1999 | Myanmar | Colombia 1979 | |
| Nigeria 1978 | St. Lucia | Comoros 1996 | |
| Romania 1991 | Dominican Rep. | Congo 1973 | |
| Seychelles 1993 | Suriname | Cuba 1976 | |
| Togo 1992 | Croatia | Dem. Rep. Congo 1978 | |
| Uruguay 1977 | Ireland | Ecuador 1995 | |
| Zambia 1996 | Uganda | Equatorial Guinea 1973 | |
| Algeria 1989 | | Finland 1999 | |
| Burkina Faso 1997 | | Guatemala 1993 | |
| Cambodia 1972 | | Morocco 1970 | |
| Cape Verde 1992 | | Sao Tome & Principe 1990 | |
| Equatorial Guinea 1991 | | Spain 1978 | |
| Fiji 1997 | | | |
| Georgia 1995 | | | |
| Guatemala 1985 | | | |
| Mali 1992 | | | |
| Mongolia 1992 | | | |
| Morocco 1996 | | | |
| Nicaragua 1987 | | | |
| Sao Tome & Principe 1982 | | | |
| Thailand 1997 | | | |
| Turkey 1982 | | | |

Table 3: *This table shows all treated and matched control states. The first column lists all states that had an increase in the actors involved in the state's appointment process. The third column lists the states that experienced a decrease in the number of actors. Columns two and four show matches for each treatment state that survived the rules for inclusion in the study and had data available for the outcome.*

the test statistic is the sum of the ranks (of the differences in time to first appointment) for all pairs such that that treated state experienced a woman appointed to the peak court first. Under this approach, a statistic of 0 would indicate that the control state experienced the appointment of a woman first for every pair. Results that are further from 0 are consistent with outcomes in which a larger number of treated states experienced the appointment of a woman first.

For causal inference, the key assumption in this design is that conditional on matching the probability of assignment to each treatment condition is equal. However, given the relatively small sample size we might be concerned that any observable difference is simply do to change. To address this possibility, we adopt a randomization approach to inference. Under this approach, the reference distribution for the statistic we calculate is constructed from our sample by permuting treatment assignments. Under the sharp null hypothesis treatment assignment is independent of potential outcomes for all units. Consequently, by permuting each treatment assignment within all pairs and then calculating the signed-rank statistic for each permutation, we can construct the distribution of test statistics under the null hypothesis.

**The Effect of an Increase or a Decrease in the Number of Actors**

Figure 7 shows the results of our procedure where we define the treatment as an increase in the total number of actors in the appointments process. The histograms display the reference distribution for our test statistic where the red, dashed line indicates the statistic calculated in the sample. The left panel shows the results for the full sample. The right panel focuses on states where our measure of the percentage of women in the national legislature is consistent with moderate pressure for diversification. In both panels the test statistic is relatively far from zero, suggesting that the treatment states more quickly experienced the appointment of a female judge than the matched, control states. Still, the one-tailed p-value is 0.38 for the full sample of states. It is 0.16 for the moderate pressure sample, precisely where we are supposed to observe the largest effect. A scholar committed to the theoretical proposition that more actors should cause diversification might be encouraged by this result, especially recognizing that the sample is still rather small (there are only 8 pairs in the moderate pressure sample). On the other hand, a fair reading of the results, especially one recognizing that a one-tailed test is likely inappropriate here, would be naturally skeptical of the finding.

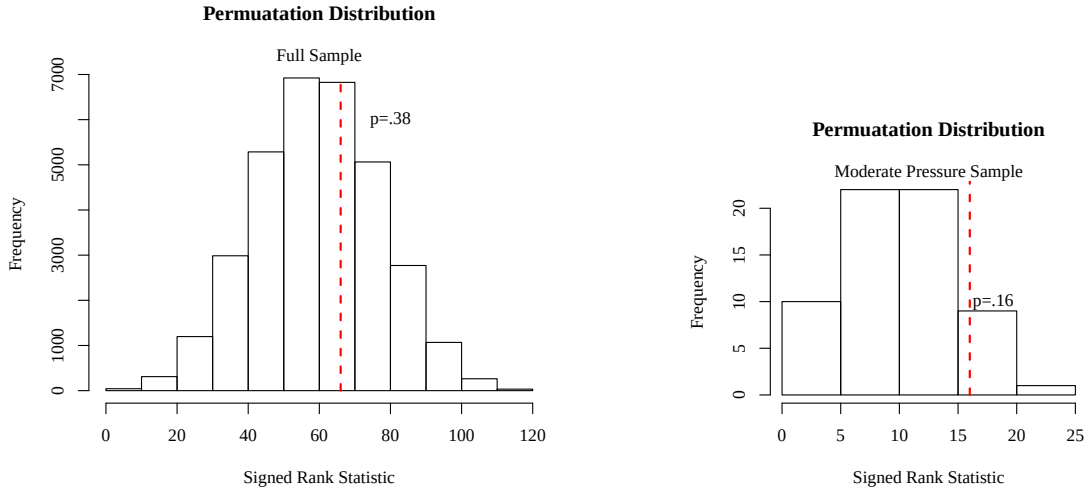


Figure 7: Increase in the Number of Actors. *Shows permutation distributions for the signed-rank statistic associated with each sample. The plot on the left shows the analysis for the full sample of treated units (15 matched pairs), whereas the figure on the right shows the same analysis for states identified as having moderate degrees of pressure for diversification (6 matched pairs). The signed-rank statistic that we calculate is displayed as the red-dashed line in the figure. One-tailed p-values are listed. For the moderate pressure test, Zambia drops out becasue it ties with Uganda, its pair.*

Figure 8 shows the results of the same analysis, but where we define treatment as a decrease in the number of actors. Here it is important to recognize that there are only five pairs of states. Again the test statistic is far from zero reflecting the fact that treatment is associated with a decreased time prior to experiencing the appointment of a woman; and, again the one tailed p-value is suggestive in such a small sample. Still, the result is consistent with precisely the opposite theoretical interpretation than what we observed in the first analysis. This is to say that here, the results suggest that reducing the number of actors involved in the appointments process should make it easier to diversify an all-male court. Thus there is evidence here for the argument that points to a small number of actors in order to avoid coordination failures.

The more important point is that, if anything, both treatments seem to cause a decrease in the time prior to the appointment of a female judge. It is worth considering what we would observe if the forces indicated by both theoretical models are at play here. In other words, what would we observe if increasing the number of actors in this process both opened up new pools of qualified candidates and created coordination problems. In this case, the effects of treatment would be attenuated toward zero. This is not what we observe. In both cases, the constitutional changes are associated with decreased time to diversification. Taken together the findings reflect a falsification of the empirical implications of each model. They are also inconsistent with a process in which the forces central to each model are at play simultaneously.

## Institutional Disruption

Our analysis began with a formalization of existing theoretical arguments, which reproduced the key logics in those approaches and yet expanded those logics in a way that identified new empirical content. This process also pointed to research design challenges, which we have addressed. Although the findings are somewhat suggestive, they are far from compelling; and critically, the pattern of findings are simply inconsistent with both of the approaches in the literature. What are we to make of this?

One natural possibility is that simple change in the the process by which judges are appointed promotes diversity, at least in the context of courts with histories of gender homogeneity. Institutional disruptions can spark high levels of elite turnover. A rich literature address how slow turnover

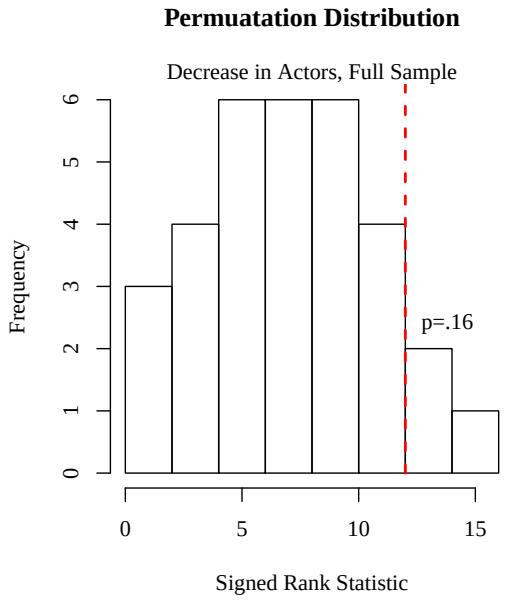


Figure 8: Decrease in the Number of Actors. *Shows the permutation distribution for the signed-rank statistic associated for sample of states in which the treated state experienced a decrease in the number of actors involved in the appointments process (5 matched pairs). The signed-rank statistic that we calculate is displayed as the red-dashed line drawn on the permutation distribution for the statistic. The one-tailed p-value is listed. All five treated units can be considered countries with moderate pressure to diversify. Albania in 1991 has approximately 3.6% women in the lower house, which is just shy of the 3.7% quartile cut off. Calculating the signed rank statistic on the four pairs excluding Albania generates a p-value of .31.*

(or high incumbency) limits the availability of contestable or open positions and subsequently dampens the accumulation of positions of power by women and other historic outsiders (Darcy and Choike 1986, Studlar and McAllister 1991, Darcy 1994, Welch and Studlar 1996, Schwindt-Bayer 2005). If an institutional change coincides with a broad re-shuffling of elites through purges, exile, or a simple change in preferences, more positions may be vacant and therefore available to women Hughes (2007).

This argument suggests that changing the institutional structure of appointing judges could increase the rate at which courts diversify. Simply changing the actors who take part in selection or the process itself, whether the result leads to greater or fewer actors, might open opportunities for diversification. This suggests an analysis in the which the treatment is defined as "any change" in the total number of actors involved in an appointments process. We exactly match each treatment state with a control state on the year of the change and by the type of constitutional event that led to the alteration of the appointment rules. These changes might happen via normal amendment procedure or when the entire constitution is replaced (as in the Nicaraguan case). This addresses the possibility that the institutional effects we identify are not actually caused by broad social changes associated with fundamental constitutional reform. Indeed, most of these changes to the judicial selection procedures occur when new constitutions are adopted, often after conflict.

Figure 9 shows the results of two analyses that pursue this line of thinking. The results in the left panel show the results of our procedure in which the treatment is defined as "any change" in the total number of actors involved in an appointments process. The results in the right panel restrict the sample to the states in which our measure indicates moderate pressure for diversification.

The results are fairly striking. The signed-rank statistics are all very far from zero and statistically significant. The finding on the full sample survives the Bonferroni correction for multiple comparisons and both findings survive the Benjamini and Hochberg (1995).

### Interpretation

We view these results as somewhat surprising in light of the institutional literature on the gender diversity of judicial appointments; however, read against the larger literature on the results of institutional change on gender diversity, what we find strikes us as fairly sensible. It would seem

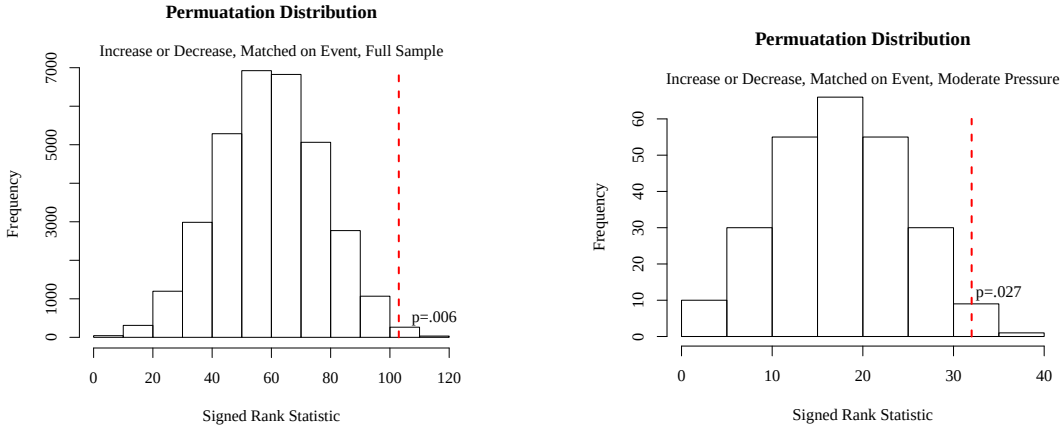


Figure 9: Any Change in Actors. *Shows permutation distributions for the signed-rank statistic associated with each sample. The plot on the left shows the analysis for the full sample of treated units (15 matched pairs), whereas the figure on the right shows the same analysis for states identified as having moderate degrees of pressure for diversification (8 matched pairs).*

that the number of actors involved in the process is unlikely to be the key driver of diversification. Rather the key institutional driver would appear to be simple institutional disruption.

Although we are confident in our empirical approach, we wish to be careful about our central claims. First, the entire study relies on formal, constitutional rules for appointments. Appointments rules defined in statutory law and informal norms of appointments lie outside of the score of the study. If these types of norms matter for gender diversity, then it is critical that we assume that they are uncorrelated with the constitutional changes we use. Second, it is important to remember that the study is set in the context of courts that have been made up of male judges up to the point of the constitutional changes in question. It remains quite possible that the theoretical forces imagined in the existing literature remain important drivers of diversification beyond the appointment of the first female judge.

Third, the countries in our study are overwhelmingly from the developing world. It is also possible that the appointments process works differently in developed states. Our key finding, that women gain a seat on courts faster after any constitutional change (largely in the developing world) is consistent with the literature that suggests post-conflict governance provides opportunities for women (Hughes 2009, Dawuni and Kang 2015). Scholarship on institutional changes resulting from violent conflict or major social/political strife suggest such an effect. This work suggests that

women make representational gains after crisis (Bauer and Britton 2006, Fallon, Swiss and Viterna 2012, Hughes and Tripp 2015, Tripp 2015).

Finally, we know that institutional rules are not fully exogenous to political life. They are chosen, typically purposively in order to resolve a social problem or manage a social phenomenon. Contracts are awarded to the lowest bidder in order to combat corruption in procurements. International agreements that lower tariffs and yet sanction particular measures to combat unfair practices are designed to eliminate inefficient trade wars. Gender quotas are adopted to increase the representation of women. As we move from theory to data, simply recognizing this general point reminds us to ask about plausible processes of sample selection in the data we observe. We have attempted to address possible sources of selection into treatment in our design. Beyond this, it is also important to consider that some rules may be viewed as exogenous, even if they were quite purposefully chosen (Carey 2000). Constitutional reform for the selection of peak court judges are very commonly adopted with political aims in mind, though by far the most common aim is to regulate the degree of independence of judges from future governments (Brinks and Blass 2018). Our study assumes that changes in appointment institutions were not adopted in order to diversify courts, or if they were, the concern was at least a second order concern.

## Conclusion

The primary goal of our study was to determine which of the competing claims in the literature about judicial selection institutions – having fewer actors involved in selection versus having more actors involved in selection – accelerated gender diversity on the bench. Reflective of the inconsistent findings in the literature, we found that an increase in the number of actors *and* a decrease in the number of actors both had positive but statistically insignificant affects on the timing of diversification. Instead, we found that any changes to the number of actors involved in selection was associated with an acceleration in diversification. This finding is consistent with scholarship that points to institutional disruptions as key causes of gender diversification in government spaces.

We believe that our study has a set of broader implications for institutional research in comparative politics. The first set of implications are methodological. Our approach is applicable in the context of any study in which scholars hope to use cross-national data on institutional changes to

draw inferences about the effects of rules on outcomes. The comparisons that are made in typical cross-national regressions rely on a series of assumptions that are unlikely to hold in most cases; however, we believe that a less often recognized consequence is that they blur our ability to see transparently what kinds of comparisons are being made across units. Our approach promotes transparency about comparisons. It is also robust to small samples, which is a likely empirical challenge for institutional scholars given the massive variation around the world in institutional structures and the infrequency of changes.

Regarding appointment institutions more specifically, the continued examination of the processes through which women come to gain seats on prestigious courts is important. We believe that future applications ought to consider more directly the bargaining environment surrounding appointments. We also envision a prominent role for studies that focus on informal norms of appointments and bring a variety of qualitative approaches to data analysis. Informal rules and norms may very well override the formal, constitutionally-based rules. We know a good deal about the *informal* appointment processes about many states, but this information has not been systematized. We believe it should be. Producing this kind of knowledge requires careful historical work, and good interviewing.

# 1 Appendix

For robustness, we now report the results of analyses in which treatment is conceptualized slightly differently. Figure 10 shows results where the treatment is an increase in the number of nominators, for both the full sample and the sample with moderate pressure for diversification. Figure 11 shows the results for an increase in the number of selectors, again for the full sample and for moderate pressure. Notably, the nominator result is statistically significant, but barely. Yet again, we see the same pattern of results in this section. We see some evidence for both of the central hypotheses of existing models, but observing both effects is simply inconsistent with either model.

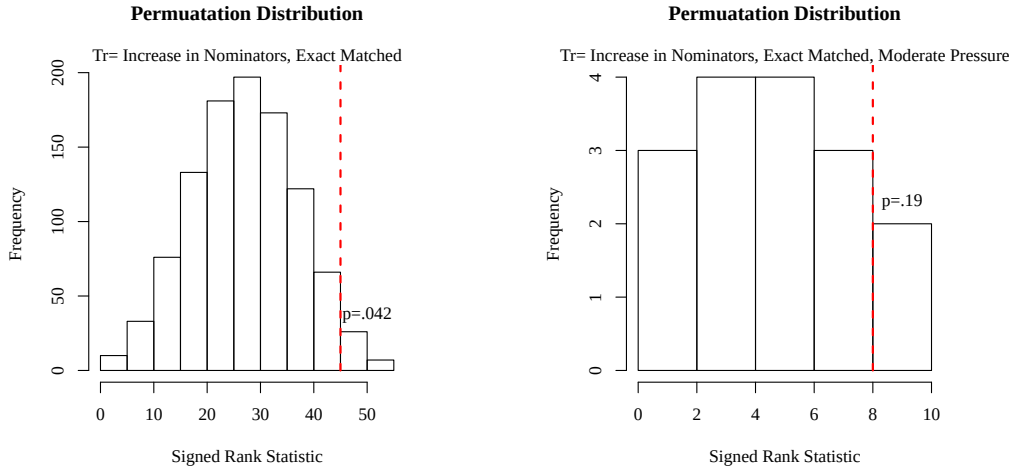


Figure 10: Increase in the Number of Nominators. *The treatment here is defined as an increase in the number of nominators. Shows permutation distributions for the signed-rank statistic associated with each sample. The plot on the left shows the analysis for the full sample of treated units (10 matched pairs), whereas the figure on the right shows the same analysis for states identified as having moderate degrees of pressure for diversification (4 matched pairs).*

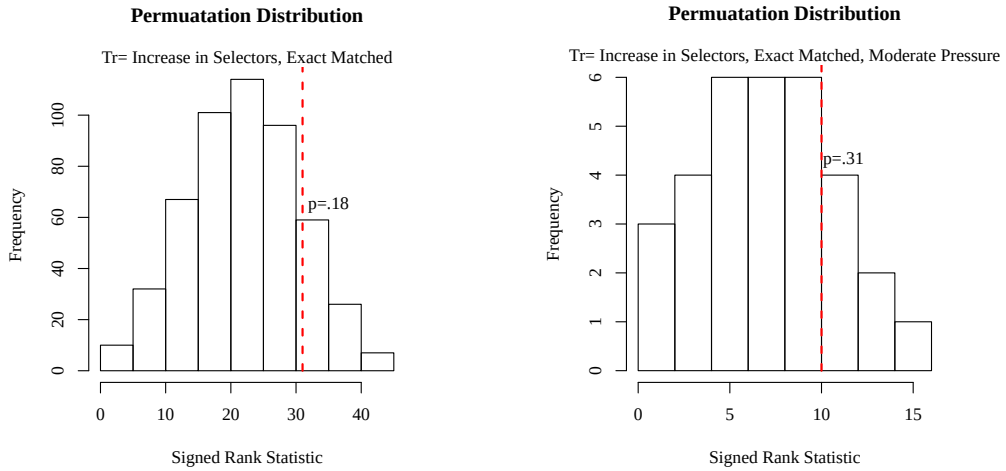


Figure 11: Increase in the Number of Actors. *The treatment here is defined as an increase in the number of selectors. Shows permutation distributions for the signed-rank statistic associated with each sample. The plot on the left shows the analysis for the full sample of treated units (9 matched pairs), whereas the figure on the right shows the same analysis for states identified as having moderate degrees of pressure for diversification (4 matched pairs).*

## References


Anleu, Sharyn Roach and Kathy Mack. 2009. "Gender, judging and job satisfaction." *Feminist Legal Studies* 17(1):79–99.

Bauer, Gretchen and Hannah Evelyn Britton. 2006. *Women in African parliaments*. Lynne Rienner Publishers Boulder, CO.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57(1):289–300.

Boyd, Christina L., Lee Epstein and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54(2):389–411.

Bratton, K. A. and Rorie L. Spill. 2002. "Existing Diversity and Judicial Selection: The Role of the Appointment Method in Establishing Gender Diversity in State Supreme Courts." *Social Science Quarterly* 83(2):504–518.

Brinks, Daniel M and Abby Blass. 2018. *The DNA of Constitutional Justice in Latin America: Politics, Governance, and Judicial Design.* Cambridge University Press.

Carey, John M. 2000. "Parchment, equilibria, and institutions." *Comparative Political Studies* 33(6-7):735–761.

Collins, Paul M., Jr., Kenneth L. Manning and Roberta A. Carp. 2010. "Gender, Critical Mass, and Judicial Decision Making." *Law and Policy* 32(2):260–281col.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Jan Toerell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kelly McMann, Daniel Pemstein, Megan Reif, Svend-Erik Skaaning, Jeffrey Staton, Eitan Tzelgov, Ti-ting Wang and Brigitte Zimmerman. 2015. "Varieties of Democracy: Codebook v4.".

Darcy, Robert. 1994. *Women, elections, and representation.* Vol. 1 U of Nebraska Press.

Darcy, Robert and James R Choike. 1986. "A formal analysis of legislative turnover: Women candidates and legislative representation." *American Journal of Political Science* pp. 237–255.

Dawuni, Josephine and Alice Kang. 2015. "Her Ladyship chief justice: the rise of female leaders in the judiciary in Africa." *Africa Today* 62(2):44–69.

Elkins, Zachary, Tom Ginsburg and James Melton. 2012. "Comparative Constitutions Project (CCP)." *URL: www. comparativeconstitutionsproject. org* .

Epstein, Lee, Jack Knight and Andrew D. Martin. 2003. "The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Cout." *California Law Review* 91:903–965.

Fallon, Kathleen M, Liam Swiss and Jocelyn Viterna. 2012. "Resolving the democracy paradox: Democratization and women's legislative representation in developing nations, 1975 to 2009." *American Sociological Review* 77(3):380–408.

Farhang, Sean and Gregory Wawro. 2004. "Institutional Dynamics on the U.S Court of Appeals: Minority Representation Under Panel Decision Making." *The Journal of Law, Economics, and Organization* 20(2).

Feenan, Dermot. 2008. "Women Judges: Gendering Judging, Justifying Diversity." *Journal of Lw and Society* 35(4):490–519.

Gill, Rebecca. 2012. "A Framework for Comparative Judicial Selection Research With an Application to Gender Diversity on High Courts.".

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.

Glynn, Adam N and Nahomi Ichino. 2015. "Using qualitative information to improve causal inference." *American Journal of Political Science* 59(4):1055–1071.

Hansen, Ben B. and Stephanie Olsen Klopfer. 2006. "Optimal full matching and related designs via network flows." *Journal of Computational and Graphical Statistics* 15(3):609–627.

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(2):216–240.

Hoekstra, Valerie. 2010. "Increasing the Gender Diversity of High Courts: A Comparative View." *Politics and Gender* 6(03):474–484.

Hoekstra, Valerie, Miki Caul Kittilson and Elizabeth Andrews Bond. 2014. *Gender, High Courts, and Ideas about Representation in Western Europe.* New York: Oxford University Press book section 3, pp. 103–117.

Hughes, Melanie M. 2007. "Windows of political opportunity: Institutional instability and gender inequality in the world's national legislatures." *International Journal of Sociology* 37(4):26–51.

Hughes, Melanie M. 2009. "Armed Conflict, Internationa Linkages, and Women's Parliamentary Representation in Developing Nations." *Social Problems* 56(1):174–204.

Hughes, Melanie M and Aili Mari Tripp. 2015. "Civil war and trajectories of change in women's political representation in Africa, 1985–2010." *Social forces* 93(4):1513–1540.

Iacus, Stefano, Gary King and Giuseppe Porro. 2009. "cem: Software for Coarsened Exact Matching." *Journal of Statistical Software* 30(1):1–27.
**URL:** *https://www.jstatsoft.org/index.php/jss/article/view/v030i09*

Kenney, Sally Jane. 2013. *Gender and justice: Why women in the judiciary really matter.* Routledge.

Kim, Haesook. 2009. "The Avalanche Perspective: Women Jurists in Korea 1952-2008." *Feminst Legal Studies* 17(61–77).

Linehan, Jan. 2001. Women and Public International Litigation. In *Background paper for the Project on International Courts and Tribunals, London Meeting.*

Malleson, Kate. 2003. "Justifying Gender Equality on the Bench: WhyDifference Won't Do." *Feminist legal studies* 11(1):1–24.

Malleson, Kate. 2014. "The case for gender quotas for appointments to the Supreme Court.".
**URL:** *http://ukscblog.com/case-gender-quotas-appointments-supreme-court/*

Moraski, Bryon J and Charles R Shipan. 1999. "The politics of Supreme Court nominations: A theory of institutional constraints and choices." *American Journal of Political Science* pp. 1069–1095.

Neto, Octavio Amorim and Gary W Cox. 1997. "Electoral institutions, cleavage structures, and the number of parties." *American Journal of Political Science* pp. 149–174.

O'Connor, Sandra Day and Kim K Azzarelli. 2011. "Sustainable development, rule of law, and the impact of women judges." *Cornell Int'l LJ* 44:3.

Osborne, Martin J. 2003. *An Introduction to Game Theory.* Oxford University Press.

Paxton, Pamela, Jennifer Green and Melanie Hughes. 2008. "Women in Parliament, 1945-2003: Cross-National Dataset.".
**URL:** *http://doi.org/10.3886/ICPSR24340.v1*

*Proportion of seats held by women in national parliaments.* N.d.
**URL:** *http://data.worldbank.org/indicator/SG.GEN.PARL.ZS?page=2*

Riker, William. 1982. *Liberalism against populism.* Waveland Pr Inc.

Rosenbaum, Paul R. 2002. *Observational studies.* Springer.

Schwindt-Bayer, Leslie A. 2005. "The incumbency disadvantage and women's election to legislative office." *Electoral Studies* 24(2):227–244.

Studlar, Donley T and Ian McAllister. 1991. "Political recruitment to the Australian legislature: toward an explanation of women's electoral disadvantages." *Western Political Quarterly* 44(2):467–485.

Tripp, Aili Mari. 2015. *Women and power in post-conflict Africa.* Cambridge University Press.

Turquet, Laura. 2011. "Progress of the worlds women 2011-2012. In pursuit of justice.".

Valdini, Melody E. and Christopher Shortell. 2016. "Women's Representation in the Highest Courts: A Comparative Analysis of the Appointment of Female Justices." *Political Research Quarterly* 69(4):865–876.

Welch, Susan and Donley T Studlar. 1996. "The opportunity structure for women's candidacies and electability in Britain and the United States." *Political Research Quarterly* 49(4):861–874.

Williams, Margaret S. and Frank C. Thames. 2008. "Women's Representation on High Courts in Advanced Industrialized Countries." *Politics and Gender* 4(03).