



Gender Diversity on High Courts

Nancy Arrington, Leeann Bass,
Adam Glynn, Jeffrey K. Staton,
Brian Delgado, Staffan I. Lindberg

October 2017

Working Paper

SERIES 2017:54

THE VARIETIES OF DEMOCRACY INSTITUTE



UNIVERSITY OF GOTHENBURG
DEPT OF POLITICAL SCIENCE

Varieties of Democracy (V-Dem) is a new approach to conceptualization and measurement of democracy. It is co-hosted by the University of Gothenburg and University of Notre Dame. With a V-Dem Institute at University of Gothenburg with almost ten staff, and a project team across the world with four Principal Investigators, fifteen Project Managers (PMs), 30+ Regional Managers, 170 Country Coordinators, Research Assistants, and 2,500 Country Experts, the V-Dem project is one of the largest ever social science research-oriented data collection programs.

Please address comments and/or queries for information to:

V-Dem Institute

Department of Political Science

University of Gothenburg

Sprängkullsgatan 19, PO Box 711

SE 40530 Gothenburg

Sweden

E-mail: contact@v-dem.net

V-Dem Working Papers are available in electronic format at www.v-dem.net.

Copyright ©2017 by authors. All rights reserved.

Gender Diversity on High Courts*

Nancy Arrington[†] Leeann Bass[‡] Adam Glynn[§] Jeffrey K. Staton[¶]
Brian Delgado^{||} Staffan I. Lindberg^{**}

*This research project was supported by Riksbankens Jubileumsfond, Grant M13-0559:1, PI: Staffan I. Lindberg, V-Dem Institute, University of Gothenburg, Sweden; by Knut and Alice Wallenberg Foundation to Wallenberg Academy Fellow Staffan I. Lindberg, Grant 2013.0166, V-Dem Institute, University of Gothenburg, Sweden; as well as by internal grants from the Vice-Chancellors office, the Dean of the College of Social Sciences, and the Department of Political Science at University of Gothenburg. We performed simulations and other computational tasks using resources provided by the Notre Dame Center for Research Computing (CRC) through the High Performance Computing section and the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre in Sweden, SNIC 2016/1-382 and 2017/1-68. We specifically acknowledge the assistance of In-Saeng Suh at CRC and Johan Raber at SNIC in facilitating our use of their respective systems.

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

[‡]Ph.D. Candidate, Department of Political Science, Emory University

[§]Associate Professor, Department of Political Science, Emory University

[¶]Associate Professor, Department of Political Science, Emory University

^{||}Department of Political Science, Emory University

^{**}Professor, Department of Political Science, University of Gothenburg

Abstract

Increasing the diversity of political institutions is believed to improve the quality of political discourse and, subsequently, the quality of political outcomes. Moreover, the presence of diverse officials in positions of power signals the openness and fairness of political institutions. These benefits of diversity should be particularly acute in the judiciary, where judges are tasked with the symbolically and substantively powerful duty of interpreting and defending constitutional values. Extant scholarship suggests that well-designed appointment process can promote diversity without explicitly gendered goals, much less quotas. If correct, these proposals raise the possibility of promoting greater diversity without having to resolve politically charged debates about quotas. Yet, scholars disagree about the effects of particular design choices. Worse, estimating causal effects of institutions in observational data is particularly difficult. We develop a research design linked to the empirical implications of existing theoretical arguments to evaluate the effect of institutional change on the gender diversity of peak courts cross-nationally. Specifically, we consider the effect of an increase (or a decrease) in the number of actors involved in the appointment process. We find mixed results for any existing claim about the role of appointment institutions play in increasing diversity. Yet we also find that any institutional change seems to cause an increase in the gender diversity of peak courts.

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 because 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. Malleon, 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 ensuring clear accountability; clear accountability facilitates the imposition of sanctions for homogenous selections and also allows for 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 coordination failures as one actor can blame another for a failure to diversify. Thus the key is concentrating appointment power so that whatever pressure there is for diversification can itself be concentrated.

Percent Women on High Courts, 2011

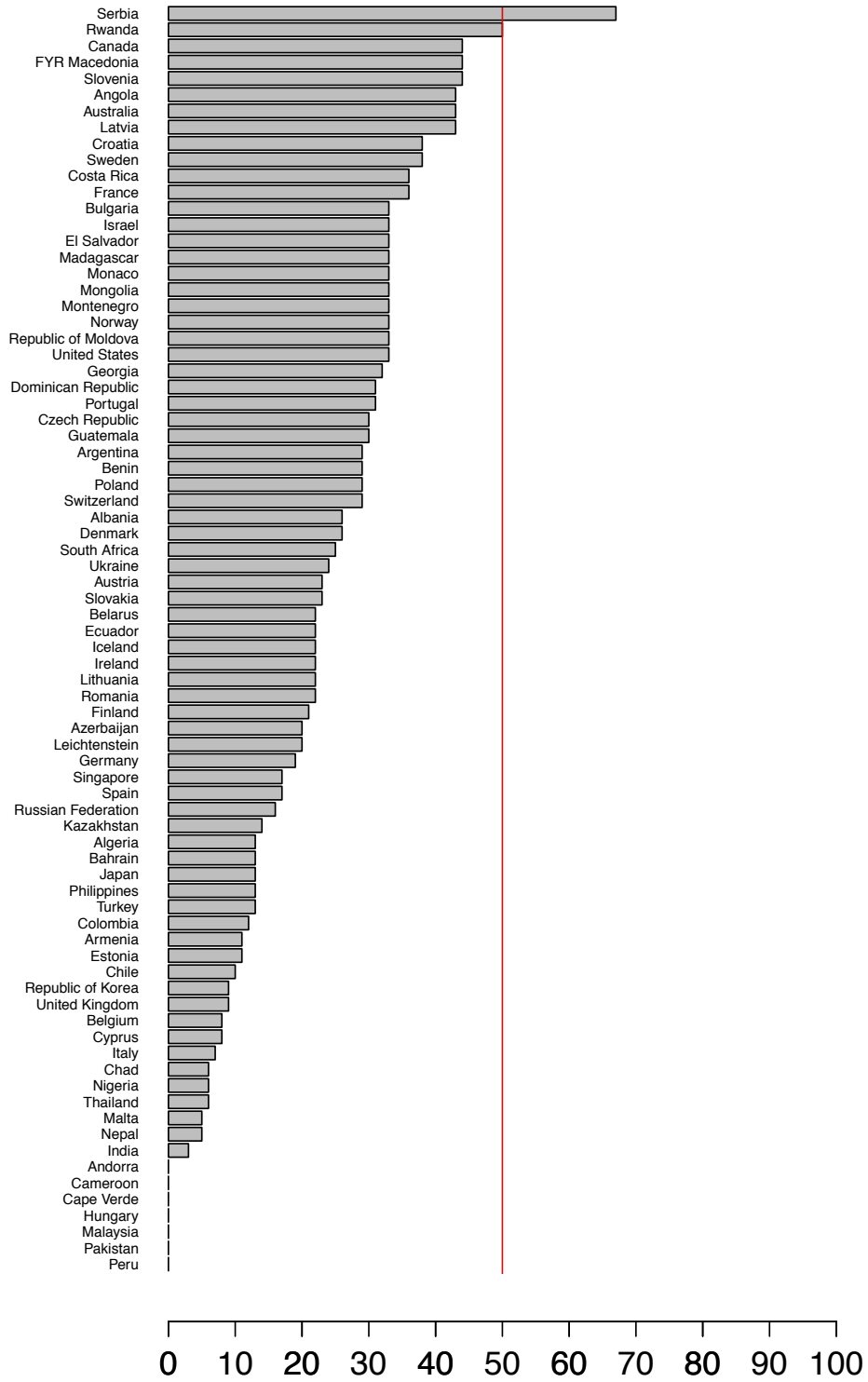


Figure 1: *Gender Diversity on Hight Courts. Data from Turquet et al 2012.*

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 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, as we do so it is important to keep an open mind about the variety of theoretical arguments and research designs that might inform the debate. We approach the problem in this paper in three ways.

Clarify Theoretical Claims It is appropriate to orient the study around existing arguments; however, we believe that the 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. Most importantly, 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).

¹Still 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.

²Specifically, 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.

Drawing on the work of Bratton and Spill and Hoekstra, Kittilson 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. In this sense, institutional effects are conditional. In addition, we will show 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.

Strengthen Research Design We will develop an empirical design that will permit credible causal claims, which are linked to our formalizations of existing arguments. 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 and related concern is that differing adoption times for appointment systems greatly complicate attempts 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. First 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

³The 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.

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 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).

Identify multiple pathways forward It may be that institutional form influences diversification in the way that one (but not the other) of the arguments we consider has identified. Yet, it is also possible that the two arguments identify forces that are at play simultaneously in any selection process. An increase in the number of actors playing a role in appointment might at once raises the possibility of a coordination failure and expand searches for qualified women judges. If this is true, then both arguments might be correct yet we would not find evidence in a cross-national study of the kind we conduct. It is also possible that the processes of appointment are captured by particular theoretical arguments well in some states but not others. This may not be a matter of two states playing different equilibria in the same game, but rather of two states operating under the rules of totally different games. And finally, it is possible that neither argument captures well the effects of institutional processes on diversification, either because informal institutional processes render the formal processes ineffective (Helmke and Levitsky, 2004), or because selection rules simply do not matter for gender diversity on courts.

Somewhat to our surprise, and yet reflecting the existing state of empirical scholarship, our analysis suggests weak evidence in favor of each theoretical argument: an increase in the number of actors involved in the selection process leads to slightly faster diversification but a decrease in the number of actors also leads to slightly faster diversification. Interestingly, when we look at all constitutional changes in the number of actors involved in judicial selection – that is, changes that result in either an increase or a decrease in the number of actors– we find strong evidence that treatment leads to more rapid diversification. This

evidence undermines the idea that distinct forces are cancelling each other out and suggests a theoretical mechanism which has yet to be identified. We summarize what that might entail; however, rather than claiming that we have a unique theoretical model to explain the mixed empirical findings, we simply point to what we believe a plausible explanation could be and highlight how a truly integrated research agenda on this subject could answer the question best. In all, we believe that our study suggests that need for a commitment to new approaches to both theory and empirics, a topic we take up in the discussion section.

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 first discuss general concerns that influence the construction of an institutional model of court diversification. Then, we give a few examples of appointment processes for peak courts and place those processes in the context of existing theoretical arguments. Next, we 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.

Assumptions about rules

We begin by recognizing that 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. Yet, recognizing this feature of institutional analysis need not prevent us from conceiving of particular institutions as exogenous for some purposes. Empirically, we may

have a sensible theoretical account of how particular rules were assigned, and via that theory, take a stab at addressing sample selection processes. Setting aside this empirical concern, simply in order to think about the likely effects of rules on behavior, it will be quite helpful to treat the rules as fixed, at least temporarily. Similarly, because conditions change over time, some rules may come to be viewed as exogenous as an empirical matter, even if they were quite purposefully chosen (Carey, 2000). The bottom line is that we will proceed theoretically by first treating appointment rules as exogenous, even though we acknowledge that the rules may be chosen in order to ensure particular kinds of appointments, potentially by the same people who are doing the appointing.

A second baseline assumption is that the types of institutions we have in mind are best thought of as “processes,” which channel individual or group interests into particular outcomes. This conception has two implications. First, appointment processes vary considerably around the world (Gill, 2012). Every theoretical statement about appointments of which we are aware ignores or obscures some element of the process in order to get theoretical traction. We will want to consider whether the elements of the process that have been set aside are important. At the very least, we should recall that our findings should be understood in the context of less than general model. Second, we should expect institutions to function differently in different places because the underlying interests that they channel vary across context. A well-known example from the study of electoral institutions makes the point. Single-member district plurality (SMDP) rules should cause greater coordination among elites and in the electorate relative to multi-member district proportional representation rules (MMPR); however, SMDP’s effect on coordination depends on underlying sociopolitical structure (Clark and Golder, 2006). In states with a high number of salient political cleavages, SMDP ought to have a powerful reductive effect on the number of candidates/parties. In states where underlying features of society have already reduced the dimensionality of political conflict, rules that incentivize coordination will have weaker (perhaps no) effects. Judicial institutions ought to exhibit this feature as well. Importantly, we see this idea at work in the Bratton and Spill study. They argue that the effectiveness of executive appointment rules for increasing gender diversity on state supreme courts turns on a court’s existing diversity. The theoretical rationale is that elections are relatively low information environments. As a consequence, voters generally are unlikely to have considerable information about existing gender imbalances on their high courts. Governors, in contrast, will surely be aware of this fact. In the context of an all-male court, governors are likely to receive considerable pressure to appoint a woman. This pressure ought to be lower once a court has been diversified. Thus, the difference between appointment rules should be particularly important for appointment choices to all-male courts (Bratton and Spill, 2002,

pp. 507-508). Generalizing this point, the effect of particular appointment rules are likely to depend on existing sociopolitical pressure for diversity in political institutions.

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 four 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. The second stage, what we will refer to as the “appointments stage” involves a legally empowered actor assigning a proposed name to a seat.

In practice, these two 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.

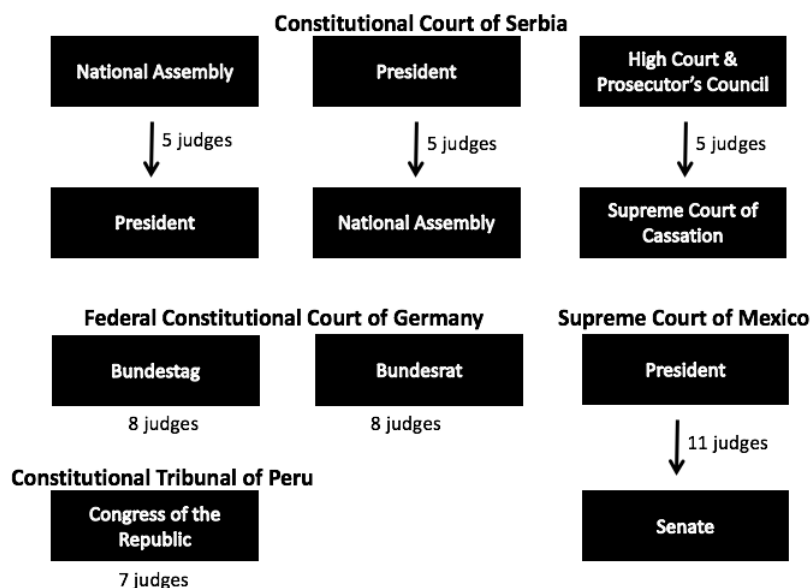


Figure 2: *Constitutional processes for appointing constitutional court judges in Serbia, Germany, Mexico and Peru. To allow a comparison to Figure 2, the processes described in the figure reflect those in practice in 2011.*

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 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.

⁴Note: 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.

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 arguments 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. Prior to developing the models, we address some general modeling concerns.

Classes of Models

Reducing appointment processes to one of two types (where a single person or institutional actor can be identified or not) greatly simplifies both theoretical and empirical analyses, but we should remember that much about existing processes has, in fact, been obscured for this benefit. First, suppose that we wish compare to compare presidential systems to all others. It is not entirely clear why the Mexican presidential system concentrates accountability more than the Peruvian system. Indeed, since both processes involve a legislative chamber, if anything, the Mexican system has a more diffuse form of accountability. Second, it is not clear why we should treat the Mexican and German systems as if they were identical to the Serbian process. Our approach will be to simply count up the actors who are involved.

Another consideration is that even within a particular institutional structure there can be considerable variation, and even just with respect to the formal rules. Consider the Mexican case in relation to another familiar process, that run in the United States. Both cases are considered forms of presidential appointment; however, they are different in a number of ways. The U.S. President nominates a single candidate. The Mexican President delivers to the Senate a slate of three nominees. In so far as a larger slate allows for a more diverse group of candidates, this difference may be quite important to the chances of appointment a female judge. Other differences might matter, as well. Whereas the U.S. President can always send the Senate another candidate independently of the number of times that the Senate has rejected a nominee, the Mexican President is empowered to select independently after the Senate has rejected two slates.⁵ If scholars are correct in their view that presidents have particularly strong incentives to appointment women (Williams and Thames, 2008), then the Mexican system, by advantaging the president, should be viewed as representing a more diversity-enhancing process.

⁵Artículo 96, Constitución Política de los Estados Unidos Mexicanos.

Of course, this entire line of reasoning suggests that we might view appointment processes for the purpose of enhancing diversity as appointment processes are generally conceived – as bargaining interactions (e.g. Moraski and Shipan, 1999). But what kind of bargaining model exactly? Without an explicit statement of the process, it is not clear what to conclude about the bargaining leverage of one party relative to another. Further, it is not clear that a bargaining framework is necessary or useful. Depending on what we believe about the differences between women and men in particular contexts, conceiving of the process differently may be useful.

Most importantly, *existing arguments do not seem to envision a bargaining setting*. Rather the focus is either on the ways that 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. 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 court. The key difference is whether choices are independent or interdependent. Our approach will be to develop a model in which choices are modeled as independent and then to model a process in which choices are interdependent. We do so by proposing a simple probability model and then a simple game theoretic model of public goods provision.

Two Models

We begin by assuming 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 $G = \{M, F\}$.⁶ A slate of appointed judges is characterized by a list $l = \{g_1, g_2, \dots, g_n\}$ of judge genders $g_i \in G$ for selectors $1, \dots, n$. We will say first that an appointment process is considered a success if at least one member of l is F . We then relax the assumption of what constitutes success.⁷ Finally, we will assume that there is pressure in society for the appointment of a woman, which we will denote $v \in (0, 1)$,

⁶We thus collapse the nominations and appointments stages. Conceptually, we view this choice as similar to assuming that selectors play an important role in the identification of quality candidates.

⁷Note: We have completed the analysis with respect to the first model. We are still working on completing the game theoretical account, as there are a number of difficult conceptual choices that have proved vexing so far.

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, designed to avoid concerns for a variety of group dynamics that motivate scholars who prefer individual appointment rules. 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	$\max(\Delta Pr(Success))$
.02	.02	.04	.07	.05
.50	.50	.875	.94	.44
.98	.98	≈ 1.0	≈ 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 $n = 1, \dots, 4$. 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.⁸ 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

⁸Note, 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.

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 female candidate with probability p_i . An 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 .⁹

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 equilibrium we require for all players that

⁹In 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.

$$v - c = v \cdot Pr(\text{at least one other appointer chooses a woman}) \quad (1)$$

$$v - c = v \cdot (1 - (1 - p)^{n-1}) \quad (2)$$

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

$$p^* = 1 - \left(\frac{c}{v}\right)^{\frac{1}{n-1}}. \quad (4)$$

There are three points to make. 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 solve the group goal. Second, Equation 3 indicates that in equilibrium, the probability that all the other $n - 1$ appointers fail to select a woman is independent of n . Instead, this probability simply reflects the ratio of the opportunity costs to the group value of adding diversity. When demand is high, the probability of this kind of failure is very low; when demand is low, the chance of a failure of this type increases up to the point where it is certain to happen (i.e., when v falls below c). The key consequence of this independence is that the probability of appointing at least one woman in this process falls monotonically with n . This probability is

$$1 - (1 - p^*)\left(\frac{c}{v}\right). \quad (5)$$

As the essential logic of the scholars who value single appointer processes suggest, 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. For very large v , note that the last term of (5) approaches zero. Thus, independently of the effect of n on the probability of one appointer's choice, the total probability of a successful search will approach 1. Likewise, as v approaches c , the second and third terms will approach 1, no

matter the value of n , so that the probability of appointing at least one woman approaches zero. With $n \geq 2$, it is at moderate values of demand where the institutional effects should be strongest.

Finally, consider what would happen if there were only one appointer, i.e. $n = 1$. In this case, everything depends on the value of v relative to c . For very high v , gender diversity will increase with certainty. For low v , gender diversity will not be increased.

What does this mean with respect to varying numbers of appointers? When v is very low, both a single appointer process and a multiple appointer process will fail to appoint a woman. For sufficiently large v , the single appointer system will produce greater diversity whereas the multiple appointer system may fail. The probability of increasing diversity is never higher with multiple appointers, and in so far as it is lower, the difference increases as appointers are added. The extent to which this difference increases depends on the underlying pressure for diversity. When this pressure is either relatively low (but sufficient high to afford any incentive for diversity) or relatively high, the increase in appointers has a relatively weak effect. When the pressure is moderate, the effect is strong.

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 having at least one woman. 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

This section suggests a number of implications that follow from formalizing existing arguments. First, whether you conceive of the process as involving essentially independent or inter-dependent 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 a single appointer increase to the process,¹⁰ as a function of pressure for diversity (v). The bottom panel shows the same results for the model in which choices are interdependent. These functions are different in a few obvious ways. First, the effects 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 discontinuity in the bottom panel reflects difference in the probability of a successful search at relatively low pressure but where a single appointer would surely appoint a woman. In that case, simply adding a single appointer implies that you will surely not appoint a woman. More importantly, both panels indicate that the largest effects (positive or negative) are at moderate levels of pressure.

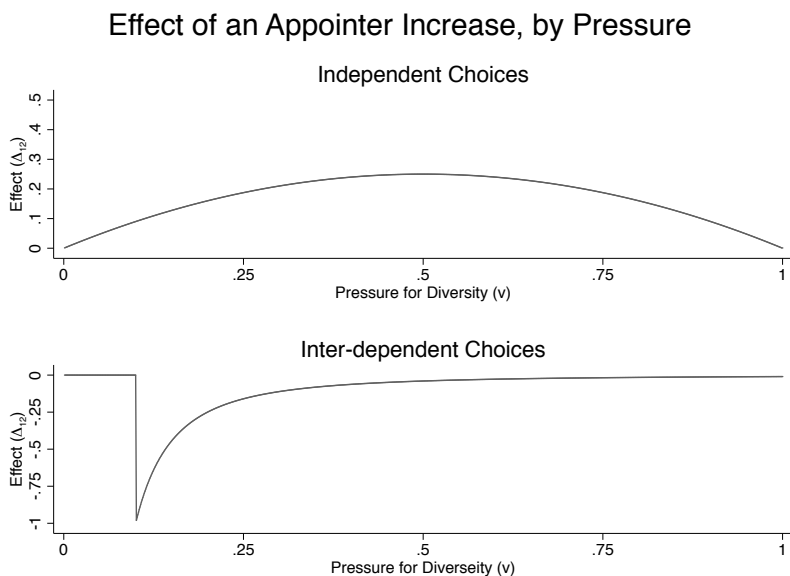


Figure 3: Shows the change in the probability of at least one female judge being appointed associated with a single appointer increase, across the complete range of pressure for diversity (v). The top panel shows results for the independent appointments model. The bottom panel shows results for the interdependent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

A final point of comparison between the models in described by Figure 4, which shows

¹⁰Specifically, 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 appointers.

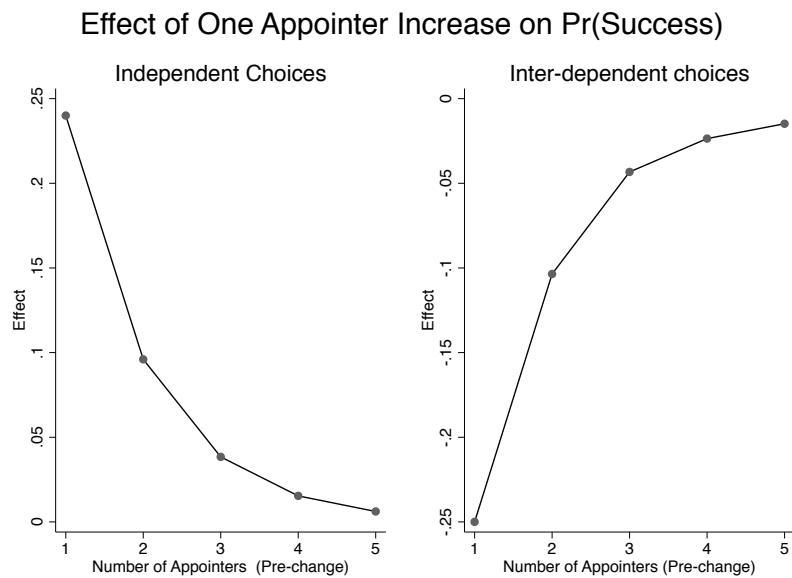


Figure 4: Shows the change in the probability of at least one female judge being appointed associated with a single appointer increase, for different numbers of appointers prior to the change. Specifically, then we consider effects associated with one appointer increases using 1, . . . , 5 as the baseline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

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 interdependent 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 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.

1. Identify states that have had changes in their constitutional terms defining the process for appointment to a peak court. We refer to these states as treatment states.
2. Exactly match each treatment state to potential control states using pre-treatment institutional information, year of constitutional change, and the number of women on the court prior to treatment. This results in a set of plausible control states for each treatment state.
3. Use propensity score matching to identify pairs of treatment and control states on the basis of plausible pre-treatment confounders.
4. For the set of states that survive this procedure, calculate a signed-rank statistic.
5. Compare this statistic to the permutation distribution of signed-rank statistics derived from our sample. The null hypothesis is that treatment assignment within pairs of

states is independent of the outcomes.

Treatment

To explore the empirical implications of the models we exploit reforms to constitutions that changed the structure guiding appointments to a state’s peak court. Again, we focus on the state’s highest ordinary court (typically but not always referred to as a “Supreme Court”) or a constitutional court if one exists. Using text provided by the Comparative Constitutions Project (CCP), we have diagrammed the appointment processes for all states included in the CCP dataset 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.¹¹ We begin by considering the effect of an increase in the number of nominators or appointers carried out between the years 1970 and 2000. For states that had more than one change in this process, we only use the first change, treating any subsequent changes in the institutional rules as post-treatment. Figure 5 illustrates this coding rule. 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. We code this as an increase from one actor in the appointment process to two actors and thus treat Nicaragua in 1987 as a treatment observation.

¹¹The 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.

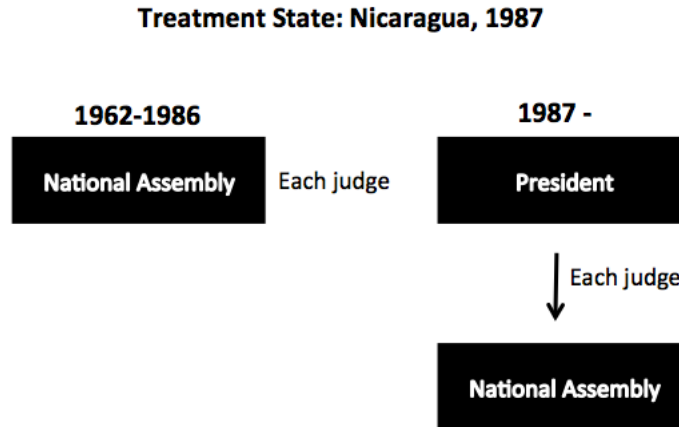


Figure 5: *Illustrates nominations and appointments process for the Supreme Court justices in Nicaragua. The process involved a single actor, the National Assembly, from 1962 to 1982. In 1987 the Constitution was reformed and assigned the nomination power to the President, leaving the appointment power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

Matching

Our interest is in making causal inferences about the effect of this treatment on a set of states that have changed their constitutional processes for appointing peak court judges and a comparable set of states that did not experience such a change. We proceed in light of our theoretical arguments. These arguments 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. In addition, we have defined success as appointing at least one women. Empirically, it is not clear whether we should treat a change from four women to five women on a seven member court as a success. 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 appropriate 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.¹² 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

¹²Data 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.

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 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.

Other included control variables from the Varieties of Democracy Project are 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.¹³

Matching Consequences 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 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

¹³We used linear interpolation to fill in any missing values in years between observed data points.

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 are matched exactly on Year, Number of Nominators, and Number of Selectors. Then within exact match subclass, treated countries are matched to one control country with propensity score matching on the other variables.*

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.

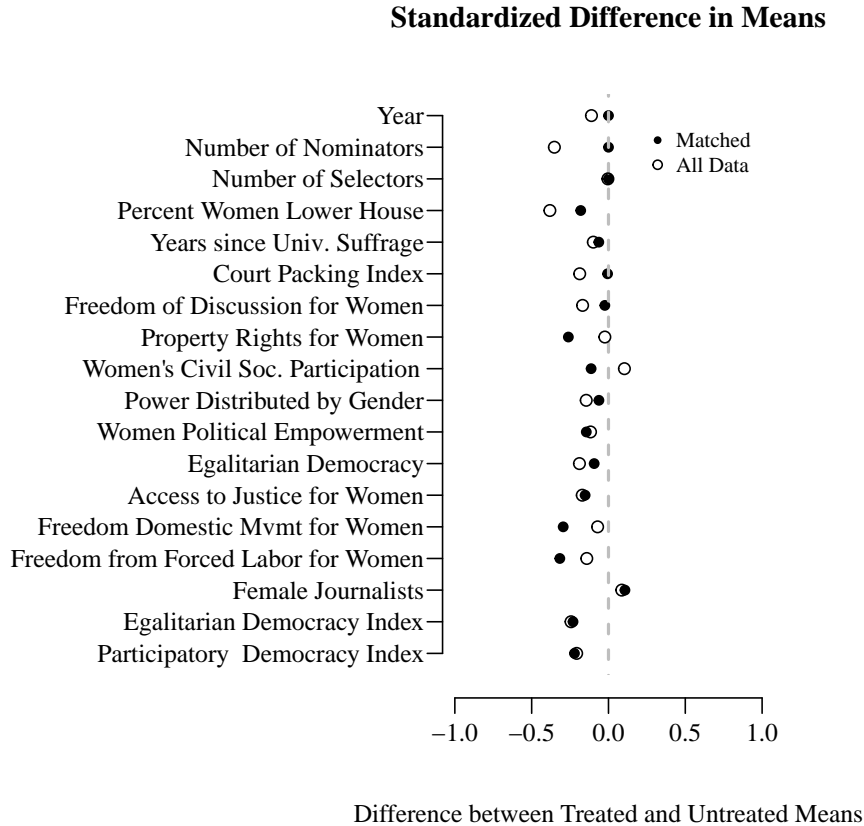


Figure 6: *Displays the standardized difference of means for treated and control units in the original and matched samples.*

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.¹⁴ 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 15 pairs of states for any analysis. Figures 7 and 8 place the treated and control units on global maps. Figure 9 shows all treated and control units if we define treatment as a change in any direction, a choice we will return to below. As Table 3 highlights treated units largely come from the developing world, where constitutions have been subject to considerable change.

Analysis: Treatment is an Increase in Actors

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.

To conduct our analysis, we first 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. The signed-rank statistic uses the sign of the difference between the time in treatment and the time in control, as well as the rank of these differences. We use the statistic as calculated in Glynn and Ichino (2015). In this approach, the 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

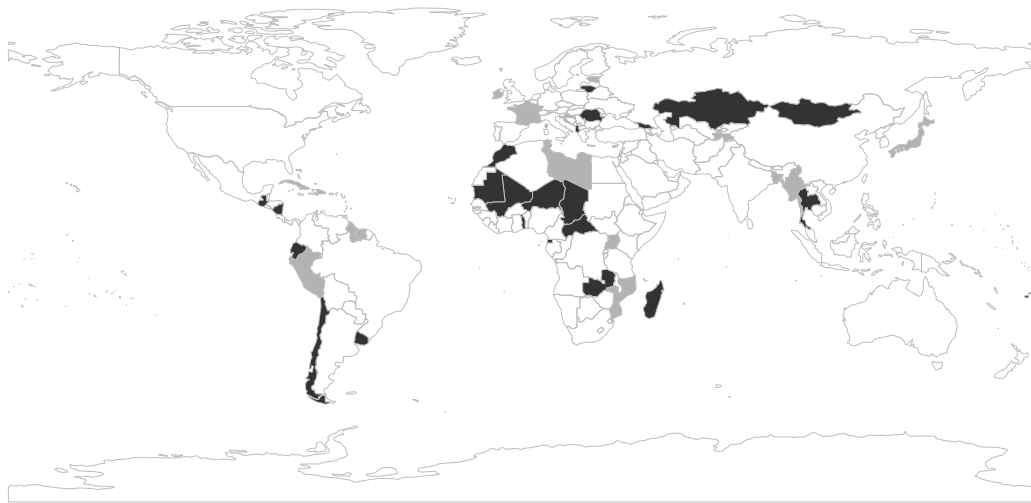
¹⁴And additional criterion is that we have data for the outcomes for each member of each pair.

<u>Treated: Increase</u>		<u>Treated: Decrease</u>	
<u>Treated</u>	<u>Control</u>	<u>Treated</u>	<u>Control</u>
Albania, 1976	Japan	Albania, 1991	Paraguay
Cntl 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	Belgium, 1998	
Lithuania, 1992	Myanmar	Bolivia, 1994	
Madagascar, 1975	France	Cambodia, 1976	
Mauritania, 1991	Estonia	Cameroon, 1975	
Nicaragua, 1987	St. Lucia	Cape Verde, 1999	
Niger, 1999	Myanmar	Chad, 1996	
Romania, 1991	Dominican Rep.	Colombia, 1979	
Seychelles, 1993	Suriname	Comoros, 1996	
Togo, 1992	Croatia	Congo, 1973	
Uruguay, 1977	Ireland	Cuba, 1976	
Zambia*, 1996	Uganda	D. R. Congo, 1978	
Algeria, 1989		Ecuador, 1995	
Angola, 1992		Equatorial Guinea, 1973	
Belarus, 1996		German Dem. Rep., 1974	
Benin, 1984		Ghana, 1982	
Bulgaria, 1991		Guatemala, 1993	
Burkina Faso, 1997		Guinea, 1990	
Cambodia, 1972		Morocco, 1970	
Cape Verde, 1992		Sao Tome & Prin., 1990	
Colombia, 1991		Spain, 1978	
Congo, 1992		Uruguay, 1985	
El Salvador, 1994			
Eqt. Guinea, 1991			
Ethiopia, 1987			
Fiji, 1997			
Georgia, 1995			
Ghana, 1979			
Guatemala, 1985			
Haiti, 1987			
Honduras, 1991			
Mali, 1992			
Mongolia, 1992			
Morocco, 1996			
Nigeria, 1978			
Poland, 1989			
Portugal, 1989			
Russia, 1991			
Sao Tome and Principe, 1982			
South Africa, 1993			
Thailand, 1997			
Turkey, 1982			
Venezuela, 1999			

Table 3: 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.

* Zambia and its pair, Uganda, drop out of the analysis because they “tie.” That is, they have the same number of years from treatment to the first woman on the court, so do not provide information for the sign-rank test.

Treated as Increase



■ Control ■ Treated

Figure 7: Displays treated and control states for an increase in the constitutional rules for nomination and appointment to a state's peak court.

Treated as Decrease

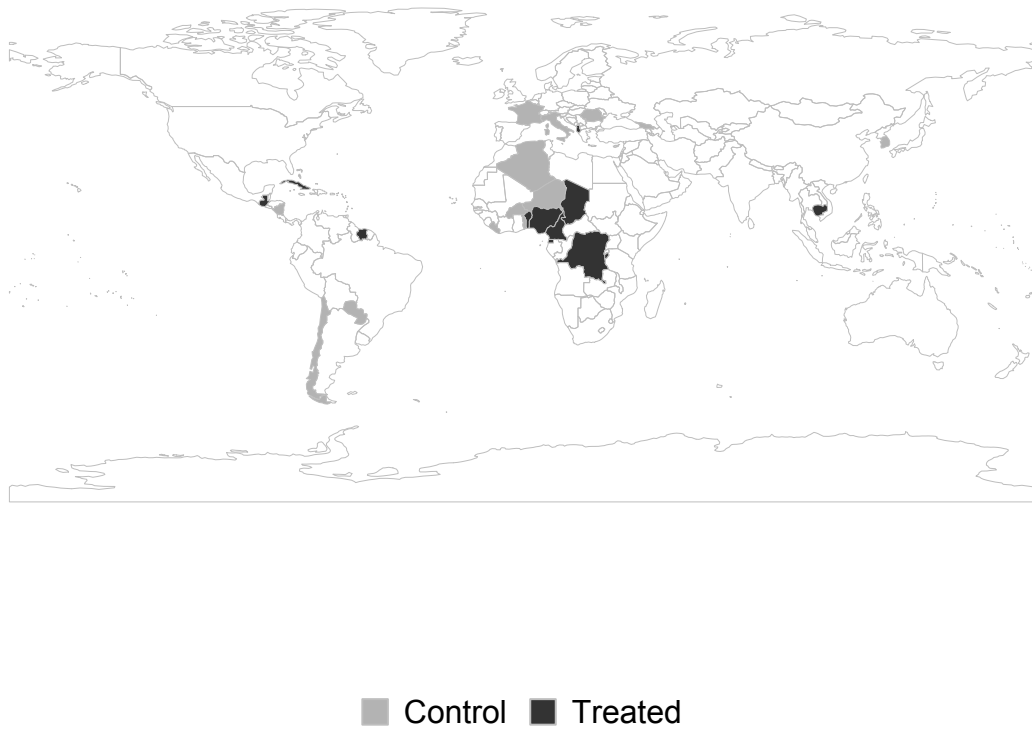
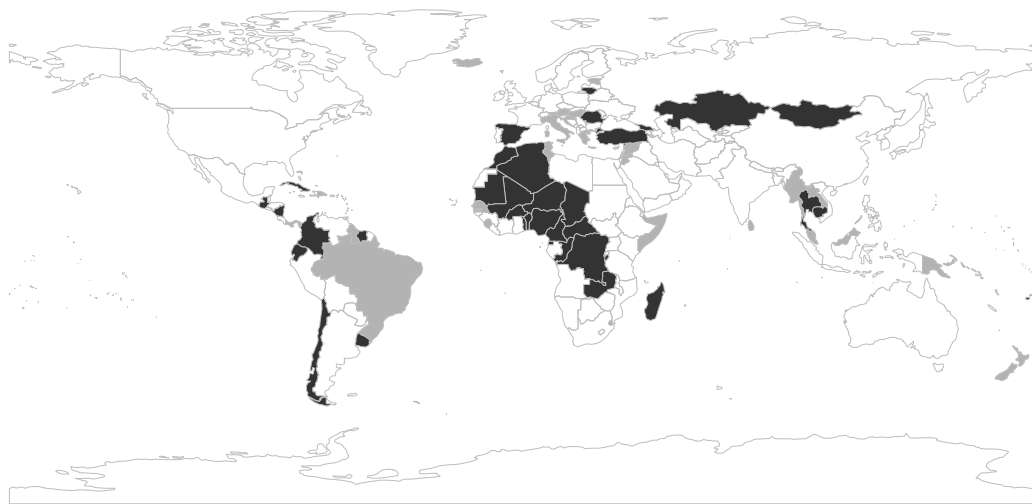


Figure 8: Displays treated and control states for an increase in the constitutional rules for nomination and appointment to a state's peak court.

Treated as All Changes



■ Control ■ Treated

Figure 9: Displays treated and control states for a treatment defined as either an increase or a decrease in the constitutional rules for nomination and appointment to a state's peak court.

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. Table 4 shows the number of pairs, p-values for each test, average years to the first woman appointee and the proportion of pairs for which the treatment observation had the first appointment for every analysis we report.

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 Fisherian randomization approach to inference. Under this approach, the reference distribution for the statistic we calculate is constructed from our sample by permuting treatment assignments. If the null hypothesis is that treatment assignment is independent of outcomes then by permuting each treatment assignment within all pairs and then calculating the signed-rank statistic for each permutation, we construct a distribution against which we can compare our result. For our smallest sample size of five pairs, there are 2^5 or 32 possible permutations of the treatment assignments. For our largest sample size of 22 pairs, there are over 4,000,000 possible permutations of the data.

Figure 10 shows the results of this procedure. The one-tailed p-value is 0.38 for the full sample of states; however, it is 0.125 in the sample that focuses on cases where the measure of percentage women in the lower chamber of legislature suggests moderate pressure for institutional diversification – precisely where we are supposed to observe the largest effect. A person 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.

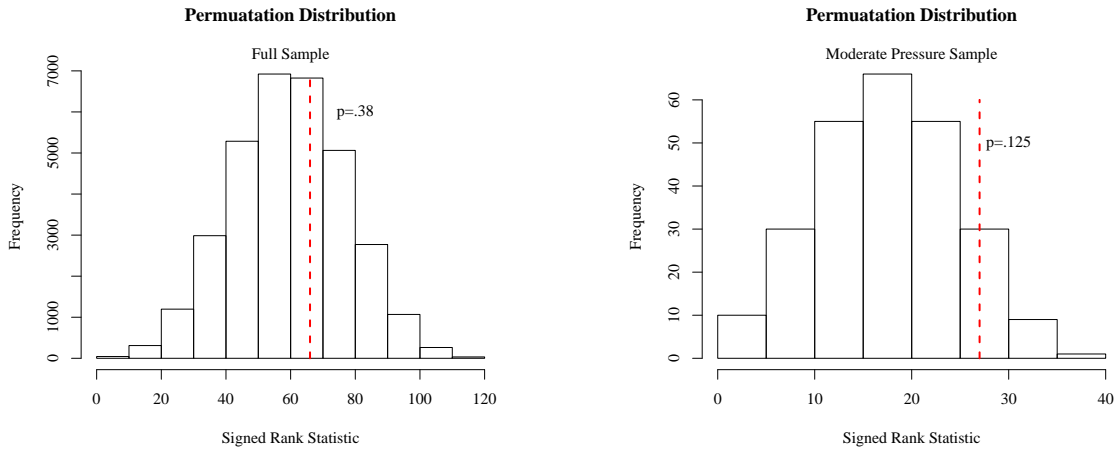


Figure 10: Shows permutation distributions for the signed-rank statistic associated with each sample. The plot on the left shows the analysis for the full sample of treated units, whereas the figure on the right shows the same analysis for states identified as having moderate degrees of pressure for diversification. The signed-rank statistic that we calculate is displayed as the red-dashed line in the figure. One-tailed p-values are listed.

Analysis: Treatment is an Decrease in Actors

Now consider the same analysis but where we define treatment as a decrease in the number of actors. As Table 3 suggests the sample size here is small as well, but we can conduct the analysis nonetheless. The p-value we obtain is correct. Again the statistic is far from zero reflecting the fact that treatment is associated with a decreased time prior to experiencing the appointment of a woman. Again, the one tailed p-value is suggestive in such a small sample. But critically, this result is precisely the opposite of what we observed in the first analysis. Both treatments seem to cause a decrease in the time prior to the appointment of a woman! These findings are consistent with what we have observed in the broader literature. They are mixed.

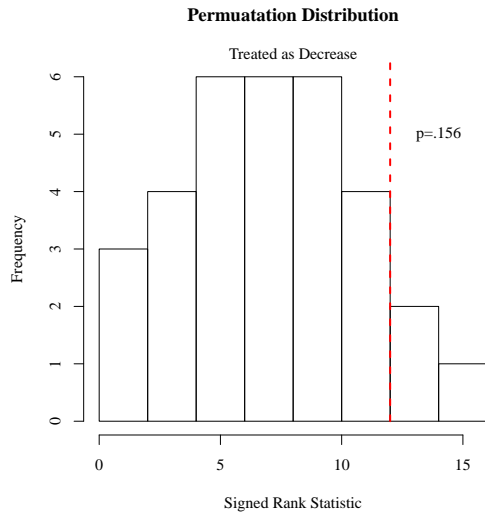


Figure 11: Shows permutation distributions for the signed-rank statistic associated for sample of states in which the treated state experienced a decrease in the number of actors involved in the appointments process. The signed-rank statistic that we calculate is displayed as the red-dashed line drawn on the permutation distribution for the statistic. One-tailed p-values are listed.

Interpretation

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 in our approach. Although the findings are suggestive, they are far from compelling; and critically, they suggest support for both arguments. Just as is true in the existing literature, it would appear that the sample you pick and the analysis you conduct powerfully influence the result that you obtain in this context. What are we to make of this?

One natural possibility is that the processes we have identified are simply not the right processes. Setting aside simple measurement error, it really may be that the formal processes that govern appointment are layered on top of informal processes of appointments that distort the incentives we see articulated in the literature. In this sense, rules matter but we have identified the incorrect ones. Another possibility is that there is not enough data in the

world to generate sufficiently powered tests. Notably our analysis of the decrease in the number of actors only involved five pairs of states. Another possibility is that the forces identified by the independent and interdependent choice models operate in different states at different times. The politics of appointments in two states are subject to different politics with vastly different incentives. And of course, the rules that we use for appointments, formal or informal, may simply not matter.

These are all sensible possibilities, albeit each suggests a quite different solution if we still wish to learn about the role that institutions play in appointments politics. Two salient alternatives remain. First, it is possible that the forces the models identify operate simultaneously. In other words, when you increase the number of actors at play in an appointments process, one simultaneously creates conditions for more diverse searches for qualified candidates and conditions for coordination failure. If that is true, the any change in any direction will result in forces being applied in opposing directions, which could plausibly result in null results under either of the two treatments we consider. But second, it is certainly possible that neither of these arguments is on the right track. It may be that some other theoretical process, very much related to the institutions, is at play. Consider the following. Suppose that we consider states that have not had a woman appointed to their peak courts. It has only been men. And suppose further that the institutional process for identifying and selecting judges has been stable. It may be that states like this with stable appointment rules will be quite likely to reproduce the gender distribution of their historical courts, for reasons associated with the ways that a variety of implicit and explicit biases manifest in hiring practices. Now suppose that the institutional process is changed, by adding or subtracting players. It may be that this change, this disruption in historical practice may lead to a reevaluation of how appointments have been done in the past. And it may be that this change causes a consideration of candidates that had historically been ignored.

We conduct one final analysis, where treatment is defined as a change in any direction – an addition or a subtraction of actors. If the theoretical forces we have considered counteract

each other, then we should not observe an effect of this treatment. On the other hand, if what matters is change, then we should see a decrease in the time between institutional change and the appointment of a woman. The results are extremely clear and highly significant. In our sample, it would appear that changes in institutional processes, really of any type, decrease the time prior to the appointment of a woman.

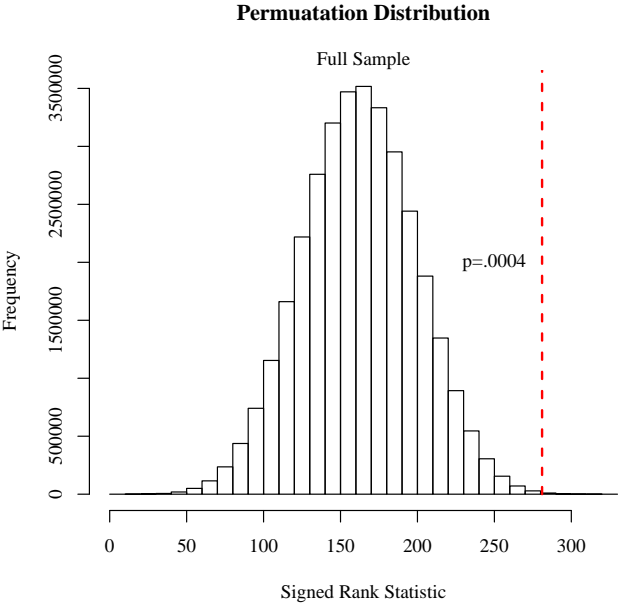


Figure 12: Shows permutation distributions for the signed-rank statistic associated for sample of states in which the treatment is defined as any change in the number of appointers or nominators. The signed-rank statistic that we calculate is displayed as the red-dashed line drawn on the permutation distribution for the statistic. One-tailed p -values are listed.

Treatment Type	Sample	Number Pairs w/ DV	p-value Signed-Rank Test	Avg. Years to First Wom., Tr	Avg. Years to First Wom., Ctrl	Prop. Pairs Tr. had Wom. First
Treatment as Increase	Full Sample	15	p=.38	11.38	13.00	.533
	Moderate Pressure	8	p=.125	10.00	17.13	.625
Treatment as Decrease	Full Sample	5	p=.156	11.00	14.90	.800
	Full Sample	22	p=.0064	10.36	17.29	.727
Treatment as Either	Moderate Pressure	17	p=.0055	9.118	15.41	.706

Table 4: Summarizes results of the permutation tests for the signed-rank statistics.

Conclusion

In this project, we set out 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. Replicating 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 effects 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. Importantly, most of these changes to the judicial selection procedures occur when new constitutions are adopted, often after conflict. Our finding that women gain a seat on the court faster after a constitutional change is consistent with literature that suggests post-conflict governance provides opportunities for women (Hughes, 2009; Dawuni and Kang, 2015).

In addition to the finding that constitutional change has a larger effect on the timing of gender diversity than the *type* of change, this project suggests a number of implications and directions for future research. First, it is possible that the the number of actors and the incentives involved in selection do affect judicial gender diversification, but that the institutional process is so complicated that casual effects are too noisy to identify. One solution to this noisiness is good experimentation to generate a controlled environment capable of revealing the casual effects of specific mechanisms.

Second, it is possible that both existing arguments (increasing or decreasing the number of selectors) are accurate but one explanation applies to some contexts and the other explanation applies to others. It may be that a public goods provision game captures the politics of state A well but not of state B. If that is correct, what is needed are country specific studies rather than cross-national studies where the analyst can come to understand the basic structure of the particular kind of politics in a state. Inference will still likely require identifying changes to rules but focusing on a single jurisdiction will mitigate the possibility of comparing two states that are simply playing different games.

Finally, it may be the case that constitutional rules are not the important rules for understanding diversification of the bench. Informal rules and norms may 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, good interviewing, and – most likely – fieldwork.

We believe that the continued examination of the processes through which women come to gain seats on prestigious courts is important. In our future work, we will consider how context – such as conflict, informal processes, and cross-country differences in how institutions condition outcomes – shape prospects for gender diversity cross-nationally.

References

- Anleu, Sharyn Roach and Kathy Mack. 2009. "Gender, judging and job satisfaction." *Feminist Legal Studies* 17(1):79–99.
- 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.
- Carey, John M. 2000. "Parchment, equilibria, and institutions." *Comparative Political Studies* 33(6-7):735–761.
- Clark, William Roberts and Matt Golder. 2006. "Rehabilitating Duvergers theory testing the mechanical and strategic modifying effects of electoral laws." *Comparative Political Studies* 39(6):679–708.
- 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."
- 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 Court." *California Law Review* 91:903–965.
- 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 Law 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.
- Helmke, Gretchen and Steven Levitsky. 2004. "Informal institutions and comparative politics: A research agenda." *Perspectives on politics* 2(04):725–740.
- 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. 2009. "Armed Conflict, International Linkages, and Women's Parliamentary Representation in Developing Nations." *Social Problems* 56(1):174–204.
- 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." *Feminist 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: Why Difference 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.
- 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.
- Williams, Margaret S. and Frank C. Thames. 2008. "Women's Representation on High Courts in Advanced Industrialized Countries." *Politics and Gender* 4(03).