

# Gender and Judicial Replacement: the Case of U.S. State Supreme Courts

Nancy B. Arrington\*

October 16, 2017

## **Abstract**

Employing a matching design on U.S. state supreme court vacancies and replacements from 1970 to 2016, I demonstrate that patterns of judicial replacement are gendered. Vacancies made by women are filled by women at a greater rate than vacancies made by men, and vacancies made by men are filled by men at a greater rate than vacancies made by women. To determine if these patterns of replacement have systematically suppressed or advantage the selection of women judges, I compare judicial selections to the gender composition of lawyers. Women are selected to state supreme courts at rates that parallel the gender diversity of lawyers over time, which suggests that gendered patterns of replacement have neither advantaged nor excluded women from state supreme courts in the aggregate.

---

\*PhD Candidate, Emory University. Contact the author at [nbarrin@emory.edu](mailto:nbarrin@emory.edu).

The importance of diversity in political office is well documented: the presence of women and minority officials in political office is associated with political engagement and participation (Junn 1997; Gay 2002; Atkeson 2003; Campbell and Wolbrecht 2006; Reingold and Harrell 2010), richer political discourse (Cramer Walsh 2002), and enhanced legislative activity and political outcomes for women and minority constituencies (Saint-Germain 1989; Phillips 1995; O’Regan 2000; Reingold 2000; Celis 2006; Reingold 2008; Wangnerud 2009). In the judicial context as well, diversity on the bench shapes perceptions of legitimacy among citizens (Scherer and Curry 2010), may alter judicial behavior and outcomes (Farhang and Wawro 2004; Boyd, Epstein and Martin 2010; O’Connor and Segal 1990), and normalizes the presence of women in positions of power (Kenney 2013).

Despite the importance of gender diversity in office, the use of gender as a selection criterion is controversial, particularly in the United States.<sup>1</sup> While many countries have turned to formal and informal gender quotas for office (Dahlerup 2008), gendered selection criteria in the US remain contentious (Baldez 2006; Krook 2006).<sup>2</sup> Opponents criticize quotas for dismantling merit selection and for favoring certain groups of people over others; they argue that descriptive characteristics should not be – and, presumably, absent these policies are not – salient selection criteria. This argument has been particularly forceful in the judicial context, where gender is viewed as inconsequential to one’s interpretation of the law: in the words of Minnesota Supreme Court Justice Jeanne Coyne, for example, “A wise old man and a wise old woman reach the same conclusion” (Margolick 1991).<sup>3</sup> It is possible, however, that descriptive characteristics are important features of the selection process even in the absence of quotas, especially if people are attentive to and there is pressure for diversification (see Goelzhauser 2011, p.776).

Indeed, studies indicate that gender is especially relevant for diversifying all-male courts (Bratton and Spill 2002), but anecdotal evidence suggests that gender matters for judicial selection on courts that have already diversified too. Specifically, there are many instances of women judges retiring and being replaced by other woman judges. If gender is irrelevant in the selection process, we would only rarely observe women judges replacing women judges: given slow turnover and the

---

<sup>1</sup>“Gender” is often conceptualized as a spectrum of identity. In this project, I use gender to refer to how individuals present in the public sphere, either male or female. For practical purposes – data availability and model simplicity – I measure gender with a binary variable.

<sup>2</sup>I use “gendered” to refer to phenomena in which “gender is present in the [process],” either implicitly or explicitly (Acker 1992, p. 567).

<sup>3</sup>Some scholarship refutes this claim that gender and experiences of men and women are inconsequential

historical dearth of women on state supreme courts, women replacing women by chance would be uncommon. In contrast, a pattern of women judges systematically replacing women judges would suggest a pattern of implicit reserved seats in which women replace women even though there is no formal rule requiring it. A pattern of women judges replacing women could, in turn, be net positive for gender diversity on the bench by ensuring minimum or status-quo levels of diversity when a court might otherwise revert to all-male or less diverse. Or, a pattern of women replacing women could suppress diversity by limiting women judicial candidates to just one or a few seats.

In this project I test whether women judges are more likely to fill vacancies made by women relative to vacancies made by men. Because broad structural forces such as the women's movement have resulted in the diversification of many professions over time and are correlated with the gender of both departing and replacement judges across the U.S. states,<sup>4</sup> I adopt a matching design to ensure credible comparisons of judicial turnover over time and across all U.S. states.

I find that when female judges retire, a much greater proportion of replacement judges are women relative to when a male judge retires, which means that gender is a relevant selection criterion for state supreme courts. To interpret this result, I compare rates of selection to the gender composition of lawyers over time. In the aggregate, women are selected to state supreme courts about as often as expected given the composition of the candidate pool. This is not to say that women judges and judicial candidates do not face implicit or explicit biases that hinder the acquisition of prestigious judicial posts or burden their experiences once in office. Instead, the evidence from this analysis suggests that at the final selection stage, women are neither systematically excluded from state supreme court benches nor unfairly advantaged.<sup>5</sup> In the next sections, I briefly summarize the extant literature, describe the research design used to identify the pattern of women judges replacing women judges, and compare those results to patterns of diversification in the candidate pool.

---

in judging. See, for example (Martin, Reynolds and Keith 2002) and (Martin 1991).

<sup>4</sup>Specifically, these broad structural forces that affect both the independent and dependent variables would invalidate an empirical strategy in which the gender of a departing judge is used simply to predict the gender of the replacement

<sup>5</sup>It is worth noting that *presence* of women judges is not the same as *equality* among men and women judges. Likewise, the absence of exclusion is not the same as the absence of discrimination. This project addresses the presence of women judges; it does not speak to the experiences of women candidates and judges.

## Consequences of and Correlates of Judicial Diversity

Scholars have and continue to debate the normative and empirical consequences of gender diversity in the judiciary. Some find that the presence of women on the bench alters judicial behavior and outcomes (Boyd, Epstein and Martin 2010; Peresie 2005; Farhang and Wawro 2004; O'Connor and Segal 1990), while others insist that these gender differences in judging have been overstated (Dixon 2009; Kenney 2008). Others emphasize the importance of gender diversity on the bench absent any gender differences or changes in outcomes. Kenney (2013) argues that the inclusion and continued presence of women on judicial benches is important in its own right because it “normalizes women’s authority and power” and demonstrates judicial legitimacy (p. 9, 175).

Scholars have also addressed the conditions under which women or minority judges are selected to the bench. Many focus on the relationship between selection procedure and diversity. Some find that the concentration of accountability on a unitary selector and the subsequent ability to credit claim is associated with greater diversity (Carbon, Houlden and Berkson 1982; Bratton and Spill 2002; Williams and Thames 2008; Valdini and Shortell 2016). Others, however, find no or little effect of selection institutions on diversity (Alozie 1988; 1990; Hoekstra, Kittilson and Bond 2014). Esterling and Andersen (1999) and Kenney and Windett (2012) highlight the importance of variation within selection procedures, noting that the diversity of merit selection committees and the gender of governors are associated with increased efforts for diversification. Other explanations for variation in judicial diversity include the diffusion of norms across space and across institutions (Williams and Thames 2008; Goelzhauser 2011; Hoekstra, Kittilson and Bond 2014); the selection of women to larger and less prestigious courts (Williams and Thames 2008); the effect of legal cultures on the acceptance of women judges (Remiche 2015); and the role of civil law systems in promoting greater gender diversity than common law systems (Schultz and Shaw 2013).

For the purpose of explaining the role of gender in judicial replacement, Bratton and Spill’s studies (2001 & 2002) represent an important point of departure. The authors consider how the existing gender diversity on state supreme courts affects prospects for the selection of women judges (2002). They find that women are most likely to be appointed to an all-male court. Once a woman is on the bench, the probability that another woman is appointed declines. In the federal judiciary, they find that President Clinton was likely to appoint minority judges to replace minority judges

(Bratton and Spill 2001). Although they could not replicate this pattern of replacement among Clinton appointees for women judges, anecdotal evidence from state supreme courts suggests that women judges often do replace women judges: when Martha Sosman left the Massachusetts court in 2007, her vacancy was filled by Justice Margot Borsford. When Justice Barbara Durham retired from the Washington state supreme court in 1999, her vacancy was filled by Justice Bobbe Bridge. In California, when justice Janice Brown was appointed to the court of appeals, she was replaced by Justice Carol Corrigan. There are many more examples, and these examples would be unexpected if gender were not a selection criterion. Only 243 of 1261 judges selected between 1970 and 2016 have been women. Given the historical rarity of women state supreme court judges, women should only replace women by chance very rarely.<sup>6</sup> This project is the first to address whether these anecdotal examples of women replacing women on state supreme court benches are systematic.

## **Pressure to Diversify and Patterns of Women Judges Replacing Women Judges**

If gender is a relevant selection criterion and women judges are being selected to replace women judges, we should consider both the causes and effects: why might judicial selectors choose or promote women judges to replace women judges? Where state supreme court judges are elected, why might women choose to run to fill vacancies by women judges? And, has the pattern of women replacing women on the bench increased the presence of women judges, as traditional quotas aim to do, or have these patterns restricted the presence of women judges by limiting women to just one or a few seats?

There are at least two explanations for a gendered pattern of replacement on courts. For one, replacing women judges with women judges could be a tool for the continued exclusion or under-representation of women. From this perspective, gendered patterns of replacement create and perpetuate tokenism on the bench. By allowing but limiting the presence of woman judges to one or a few specific seats, “tokenism is ... symbolic equality” (Greene 2013, p.82) that outwardly demonstrates a commitment to equality without addressing the underlying social and political treatment of historically marginalized groups (Laws 1975, p.51). By limiting women judicial can-

---

<sup>6</sup>To illustrate, women have been 19% of judges selected since 1970. Assuming (for simplicity sake) that women comprise 20% of current judges and 20% of judicial candidates, if gender were irrelevant in selection, on average, only 2.5% of vacancies would be by women judges who are replaced by women judges.

didates to vacancies made by women candidates, a token seat or seats excludes women from other vacancies and limits diversity on the bench.

Alternatively, patterns of women replacing women may promote diversity on the bench. Rather than limiting women to one or a few seats, gendered patterns of replacement may ensure diversity by promoting the selection of women when courts might otherwise revert back to all-male or less diverse. This explanation for gendered replacement would be particularly beneficial for diversity when there are few women in the traditional candidate pools. The historical exclusion of women and minority individuals from higher education and posts that serve as informal qualifications for judicial office has limited and continues to limit the diversity of the candidate pool. In this setting of limited availability, patterns of gendered replacement might encourage judicial selectors to seek out those women judges who *are* qualified when they might otherwise select male judges.

For seats in which judges are elected, the same forces can be at play. If party officials and donors only seek out and support women candidates for judicial races to fill vacancies by women – or if they discourage women from running for vacancies made by men – they could depress the presence of women on the bench. In contrast, if party elites and donors actively seek out women to run for a vacancy made by a woman when another woman might not run otherwise, they would be promoting diversity on the bench.

From the outset, it is unclear if gendered patterns of replacement have been beneficial for overall levels of diversity or not. Even if patterns of gendered replacement have been net positive for gender diversity on state supreme courts, these patterns may still have some pernicious outcomes for women (or minority) judges individually. For example, these patterns of women replacing women are consistent with concepts of “tracking” in which “indirect bias” funnels women and minority officials into particular posts or types of posts (Reingold and Smith 2017). Women judges selected to replace women judges may also be subject to pressures of tokenism or to backlash for perceived gender-based favoritism. It is also important to note that the use of gender as a selection criterion can be explicit or implicit. Judicial selectors may actively and knowingly seek out women judges to replace women judges, but they may also do so unintentionally. Uncovering the internal motivations of judicial selectors and the experiences of individual women judges on the bench is beyond the scope of this project. Instead, I test whether patterns of gendered replacement are systematic

---

<sup>7</sup>Given the same candidate pool.

and whether those patterns have suppressed overall gender diversity on the bench relative to a counterfactual in which gender is not a selection criterion.<sup>7</sup>

## Data and Matching

Assessing whether the gender of vacating and replacement judges are independent presents a few methodological challenges. First, the gender of the retiring judge is by no means the *only* explanation for the gender of the replacement judge, so we must isolate the effect of gender, all else equal. Second, more women have been able to accumulate the qualifications for office over time. As the feminist movement took hold and played out, citizens came to accept women on courts and – for some – expect women on courts.<sup>8</sup> These over-time pressures mean that women are more likely to be selected to a vacancy over time, regardless of the gender of the retiring judge. To accurately identify the effect of gender, we need to account for time trends. Finally, to determine the effect of the gender of the vacating judge, there must be vacating judges *who are female*. Assessing the effect of the gender of the vacating judge has only recently been possible as more women judges have been selected to and have left state supreme court benches, so we need to manage inconsistent data availability over time.

To address these methodological concerns, I do two things. First, I use non-parametric matching to generate a data set of treatment (a woman retires) and control (a man retires) cases that share theoretically and empirically important characteristics. This allows me to isolate the effect of the gender of retiring judges on the gender of replacement judges by comparing the outcome across the treatment and control.<sup>9</sup> I match on time, the number of women, court size, the number of vacancies, and selection institutions to ensure plausible comparisons across treatment and control units. Second, I use a Cochran-Mantel-Haenszel (CMH) test of proportions to determine if the gender of the retiring judge is independent of the gender of the replacement judge. The CHM test is flexible to homogenous effects across time and other covariates and can accommodate differences

---

<sup>8</sup>See, for example, Andrews (2016) and Bakalar (2016), two newspaper opinion pieces advocating the selection of a woman judge to the Alaska Supreme Court.

<sup>9</sup>The ideal way to identify and isolate the causal effect of a vacating judge’s gender on the gender of the replacement judge would be to randomly assign gender to judges. Then, we could compare the gender of vacating judges to the gender of replacement judges knowing that all other traits are equal. Clearly this is not a feasible design. Instead, matching observations increases the balance across cases in which women and men retire, which helps isolate the causal effect of vacating judge’s gender on replacement gender (see Boyd, Epstein and Martin (2010) for a more detailed discussion of matching and causal inference).

in sample size across strata, which occurs in the data because more women have vacated the bench in recent years. Before describing the results of the CMH test, I detail the data and the matching procedures used.

**Data** The gender of judges retiring from and selected to state supreme courts comes from Kathleen Bratton’s State High Court and Justice Database.<sup>10</sup> The data set includes judges selected to all 50 state supreme courts between 1960 and 2010 and describes how justices were selected, when they were selected, when they retired or vacated the court, and their gender.<sup>11</sup> I updated the data set to include judges who retired or were selected between 2010 and 2016 with information from Ballotopedia.<sup>12</sup> I restrict the data to the years between 1970 and 2016 to avoid data missingness in the early years.

The data are re-shaped so that the unit of analysis is state-years for which a judge vacates. While it is most common for a court to have only one vacancy at a time, there are many courts and years with multiple vacancies (see table 1). Aggregating to state-year rather than treating each vacancy as the unit of analysis avoids an independence problem in cases with multiple vacancies: when two or more judges are replaced at the same time, the characteristics of one replacement judge might affect the probability the other replacement judge holds certain characteristics as well. Furthermore, evidence from the legislative arena suggests that the selection of multiple candidates at once, such as on a party list, is associated with increased gender diversity (Salmond 2006; Paxton 1997; Kenworthy and Malami 1999). By aggregating to state-year and then matching on the number of vacancies, I avoid potential bias from the interdependence of vacancies and I control for potential incentives to select women when there are multiple vacancies.

I link vacancies to replacements by time. If a vacancy and a replacement occur in the same year, those two judges are linked as a vacancy/replacement pair. Importantly, retirements and replacements do not always occur in the same calendar year. For example, a judge may retire

---

<sup>10</sup><http://www.lsu.edu/faculty/bratton/research.htm>

<sup>11</sup>Texas and Oklahoma each have two high courts, one for civil cases and one for criminal cases. The criminal courts are not included here.

<sup>12</sup>Of the 1407 judges selected between 1960 and 2016 in this combined data set, there are 243 judges identified as women; 85 identified as Black or African American; 28 identified as Latinx, and 19 identified as Asian-American or Pacific Islander. Of the 243 women judges, 21 are Black women, 8 are Latina, and 8 are Asian-American/Pacific Islander women. Indicator variables for gender and Black judges are included in the State High Court and Justice Database. Indicator variables for Latinx and Asian American/Pacific Islander judges were generated through searches of NALEO directories and the Asian American and Pacific Islander



Table 1: Frequencies: Treatment and Dependent Variable

	Male Only	Female	Total
Vacating Judge (TV)	562	109	= 671
Replacement Judge (DV)	476	195	= 671

*The gender of retiring and replacement judges in the cleaned but un-matched data set. The majority of retiring and replacement judges are male. There are 109 “treated” units in which a woman judge vacates.*

in one year but her replacement may not be selected until the next year. In these cases when judges vacate or are selected in different years, I aggregate across two years to link the vacancy and replacement (see the Appendix for a description of the coding rules for aggregating two years). After cleaning and aggregating the data, I am left with 671 units that correspond to state-year(s)-vacancy(s) observations.<sup>13</sup>

The main independent variable (the “treatment” variable) is a dummy variable indicating if the vacating judge is female, and the dependent variable is a dummy variable indicating the gender of the judge selected to fill the vacancy. In cases where there is more than one vacancy and replacement, the dummy variable indicates if any of the vacating or replacing judges are female.<sup>14</sup> Table 1 summarizes the frequency of the treatment and dependent variables for the 671 units prior to matching. There are 109 “treated” cases in which a woman vacates the bench. In the next section I describe how I match those treated cases to control cases and then test for a relationship between the gender of the vacating judge and the gender of the replacement judge.

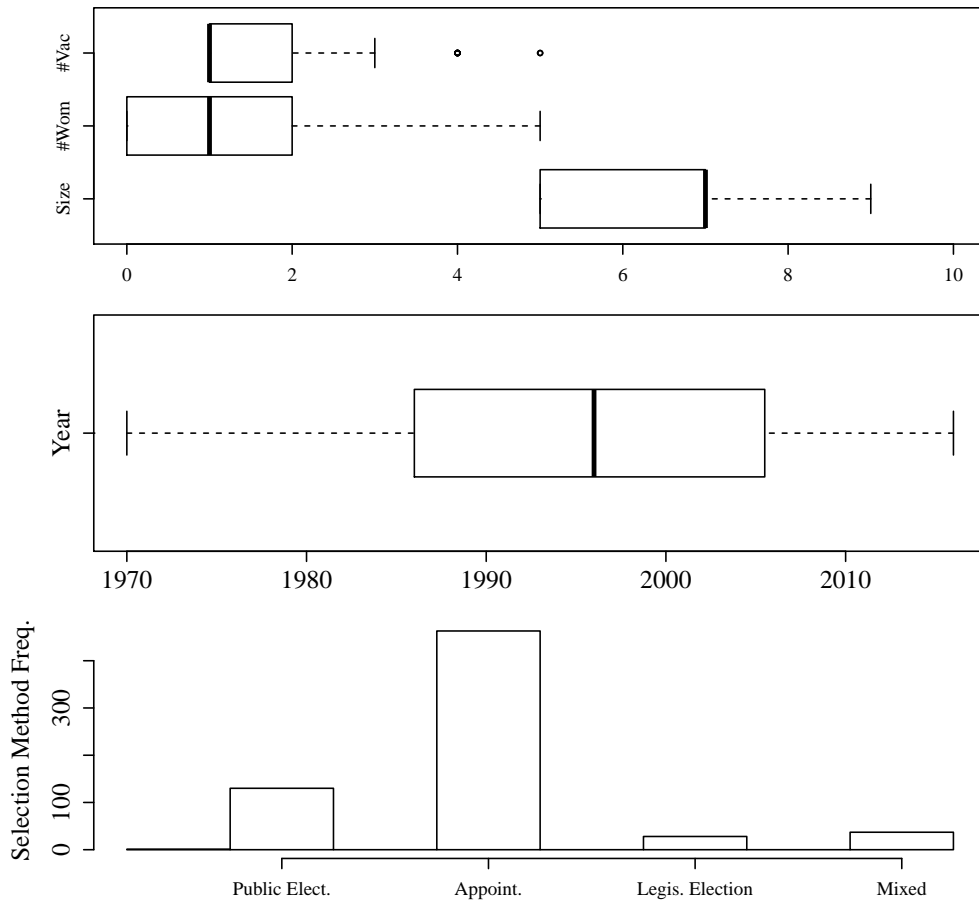
**Matching** In order to better approximate an experimental study, I employ a matching design to minimize imbalance, which in turn reduces model dependence and bias (King and Zeng 2006; Iacus,

Almanacs. In addition, keyword searches of Asian American, Pacific Islander, and Latino/a judges were used to augment missing directory and almanac years. Finally, when photographs of judges suggested a racial or ethnic identification, electronic newspaper searches or biographies were used to confirm race or ethnicity.

<sup>13</sup>To avoid counting a judge as his or her own replacement, I remove 44 judges who were selected and removed in the same year. When possible, I connect these brief-tenure judges’ predecessors to their replacements; that is, I just skip over judges who served for less than a year. Additionally, there are 107 state-year-vacancies in which the number of vacancies and the number of replacements do not match up. These observations do not fit the criteria for two-year aggregation described in the appendix. For 49 of these observations, there is a vacancy but no corresponding appointment. For 58 of these observations there is at least one vacancy and one replacement, but the number of vacancies and replacements do not correspond. I remove these observations to prevent incorrectly attributing replacements to vacancies.

<sup>14</sup>To be clear, I match observations on the number of vacancies, so courts with two vacancies are only

Figure 1: Data Summary, Unmatched Data



*The top panel shows distributions for the number of vacancies, the number women on the court, and the court size for the full, unmatched data-set. The middle panel shows the distribution of observations across time, and the bottom panel shows the distribution of selection methods. For the selection method, “public election” refers to both partisan and non-partisan elections. “Appointment” refers to gubernatorial appointment and merit selection, and “mixed” refers to years in which multiple judges were selected and the judges were selected in different ways.*

King and Porro 2011). Cases are matched based on characteristics that affect the probability that the replacement and/or retirement judge is female. I match treatment and control cases on the (1) number of women on the court, (2) the size of the court, (3) time, (4) the number of vacancies, and (5) the selection method.

Matching on the number of women on the court is important for two reasons. First, the greater the number of women on the bench the greater the probability that any given vacancy is made by a woman. Second, as Bratton and Spill (2002) show, the number of women on the bench is negatively associated with the probability a woman is selected. Importantly, because it is only possible for a woman judge to retire from a bench on which there is at least one woman, all matched data have at least one woman on the bench.<sup>15</sup>

The size of the court is an important matching variable for three reasons. First, court size affects the probability that there is a vacancy. All else equal, the more judges there are on a bench the greater the opportunities for a vacancy. Second, the court size affects our interpretation of gender diversity. One additional woman has a greater effect on the gender composition of a 5 person court than a 9 person court. Third, extant literature suggests that women are more likely to be selected to larger courts (Williams and Thames 2008; Cook 1984).

Matching on time controls for the relationship between the presence of women on courts over time. More women have been selected to and have retired from state supreme courts in more recent years. In addition, matching on time controls for variation in pressure to diversify courts over time.

Matching on the number of vacancies is important for two reasons. First, the probability a woman vacates or is selected increases in the number of vacancies. Second, if the incentives for selecting women candidates in the legislative context apply to the judicial context, women may be more likely to be selected when there are multiple vacancies (Salmond 2006; Paxton 1997; Kenworthy and Malami 1999).

Finally, observations are matched on selection method. Selection methods are grouped into three categories: popular election (both partisan and non-partisan), selection by elites (gubernatorial selection or merit selection), and legislative election.<sup>16</sup> While these groupings of selection procedures

---

compared to other courts with two vacancies.

<sup>15</sup>This count includes the gender of the vacating judge.

<sup>16</sup>In cases in which there were multiple vacancies and the replacement judges were selected through different procedures, the selection process was classified as “mixed,” however none of these were successfully matched, which is not surprising given the few observations with mixed selections.

are broad and obscure variation within groupings, these categories capture important variation in opportunities for accountability over judicial selections and in the amount of coordination required to select judges. In popular election systems, accountability for the composition of the court is very diffuse, which means that any sanctions for perceived exclusion will be diluted, and credit-claiming for diversifying (Valdini and Shortell 2016) will be less lucrative. In addition, coordination requirements to mobilize for the selection of a woman judge are high in the electoral context. In contrast, when only one or a few elites choose judges, accountability for a homogenous court is more easily attributed to those responsible and fewer people must coordinate in order to intentionally select a woman judge. Matching on these broadly-defined selection methods ensures that general patterns of accountability and coordination are held constant across treated and control groups without seriously restricting the ability to successfully match. Figure 1 summarizes the matching variables for the full, unmatched data set.

Treatment and control cases are matched using Coarsened Exact Matching procedures (Iacus, King and Porro 2011) in the “MatchIt” package in R. Observations are *exact* matched on the court size, the number of women on the bench, and the three-category selection method. For these variables, the difference between a value of one and two, for example, is substantively different, so treated cases should only be matched to control cases that share the exact values for those variables.

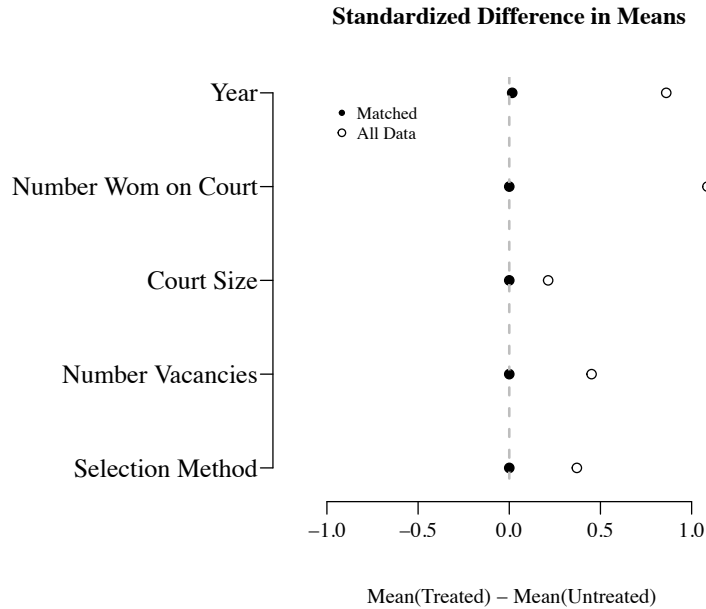
For the year variable, cases should be matched within the same social and political context but need not be matched in the exact year. The difference between a vacancy in 1994, for example, and 1995 is not substantively meaningful. Moreover, vacancies on state supreme courts are relatively rare: only about half of the states in any given year have a vacancy on the court. Matching exactly on year seriously restricts the number of matched pairs. For these two reasons, the year variable bins are coarsened according the CEM coarsening algorithm, which matches treatment cases to control cases within five-year spans.

Figure 2 shows a balance plot that summarizes how treated units compare to control units, both for the full, unmatched data set (open circles) and for the matched data set (black dots).<sup>17</sup> The farther from zero the standardized difference of means is, the greater the difference between treated and untreated observations. There are substantial differences across the treated and untreated

---

<sup>17</sup>The standardized difference in means are:  $[(\text{mean of treated units} - \text{mean of untreated units}) / \text{the standard deviation}]$ . The standardized means of matched units are calculated with the CEM weights.

Figure 2: Balance Plot



*This plot shows the standardized difference in means between the treated and untreated data for the full data set and for the matched data set. because each treatment case can be matched to multiple control cases, standardized differences in means are weighted.*

groups in the unmatched data: treated units occur later (greater value for year); there are slight differences in the frequency of selection method; treated units have more women on the bench, have more vacancies, and have a larger number of seats. If treatment and control units are matched appropriately, balance should improve and the differences in means of the matched data should be closer to zero than the differences for the full data set. As the black dots in figure 2 show, balance is greatly improved.

## Results

The Coarsened Exact Matching procedure groups treated and control cases into strata; each strata contains at least one treated unit and all the matched control units, of which there can be many. Table 2 shows a contingency table for the treatment and outcome variables for the matched data across all strata. The rows correspond to the gender of the vacating judges, and the columns show the proportion and number of vacancies filled by male and female judges. In this matched data set, when a male judge retired, more than three quarters of replacement judges were likewise men.

Only 24.8% of vacancies made by male judges were replaced by female judges. In contrast, when a woman judge retired, almost half (44.8%) of her replacements were also women.

Table 2: Matched Data, Contingency Table

		Replacement Judge		
		Male	Female	Total
Vacating Judge	Male	.75 (100)	.25 (33)	=1 (133)
	Female	.55 (32)	.45 (26)	= 1 (58)
C-M-H		$\chi^2=6.8$ p=.01		

*This contingency table summarizes how many vacancies were made by men and women and how many of those vacancies were filled by men and women. Proportions and total number of cases (in parentheses) are reported. The Cochran-Mantel-Haenszel test (C-M-H) generates a Chi-squared statistic of 6.8 with a corresponding p-value of .009, which means that we can reject the null hypothesis that the gender of the replacement judges is independent of the gender of the vacating judge.*

To determine if the proportions of female judges selected as replacements are sufficiently different across male and female vacancies, I use a Cochran - Mantel - Haenszel test (CMH).<sup>18</sup> The null hypothesis of the CMH test is that there is no association between treatment and outcome variables. Under the null hypothesis, the proportion of women selected to fill a vacancy is independent of the gender of the vacating judge. For these data, the CMH test produces a  $\chi^2$  statistic of 6.8 with one degree of freedom.<sup>19</sup> The proportions of male and female judges selected to fill vacancies made by men and made by women are sufficiently different to reject the null hypothesis that the gender of the replacement judge is independent of the gender of the vacating judge (p-value=.009). This test indicates that the gender of the retiring judge does affect the gender of the judge selected to fill the vacancy. Vacancies made by women judges are filled by women judges at a greater rate

<sup>18</sup>The CMH test is appropriate for binary treatment and outcome characteristics across differently sized strata (McDonald 2014, p94-100).

<sup>19</sup>Calculated with the 'mantelhaen.test' package in r.

than vacancies made by men, and vacancies made by men are filled by men at a greater rate than vacancies made by women.

While this analysis demonstrates that gender is relevant in the selection process, it does not distinguish between a positive or negative outcome for judicial diversity overall. In the next section, I compare patterns of judicial selection to the gender composition of lawyers to determine if patterns of gendered replacement have suppressed or promoted the selection of women judges to state supreme courts.

### Gender, Judicial Replacement, and Diversification

If the pattern of women judges replacing women judges requires judicial appointers or party elites to actively seek out women candidates when they otherwise would not, then this pattern may promote gender diversity. In contrast, if the pattern of women replacing women *limits* opportunities for women judges by restricting them to one or a few seats, then this pattern could suppress diversification. To investigate whether current patterns of selection promote or suppress the selection of women to state supreme courts, I compare the observed rates of selection to a counterfactual in which patterns of selection do not depend on the gender of vacating judges.<sup>20</sup> I use the gender composition of lawyers as a proxy for the candidate pool for state supreme court judges.<sup>21</sup> Then, I estimate the rate at which women *would* be selected to state supreme courts if gender were not a relevant selection criterion under existing and broadly defined standards of what it means to be qualified.

It is important to note that there is no agreed upon, ideal level of gender diversity on state supreme courts. In some contexts, such as the legislative context, standards of descriptive representation and diversity are compared to the composition of the population. The assumption, sometimes implicit, is that if political institutions are open and available to all, then the characteristics of the representatives should generally mirror the characteristics of the population. In the context of high courts, though, the vast majority of the electorate is not formally *eligible* for office

---

<sup>20</sup>I am not estimating what the rate of selection of women judges *should* or *ought* to be. For a discussion about the *right* or *just* rate of selection of women judges see, for example, Kenney (2013) or Mansbridge (1999).

<sup>21</sup>Cook (1978; 1984); Alozie (1996); Kenney (2012); Solberg and Bratton (2005) and others have also used the gender composition of lawyers as an estimate of or variable affecting the candidate pool for judicial posts.

under existing rules. At least 38 of the 50 states, for example, require state supreme court judges to be lawyers.<sup>22</sup> Requirements include practicing law for a certain number of years, being a member of the state bar, being a licensed attorney, or “being learned in the law.”

Given how many states require judges to be lawyers, I choose the gender composition of the lawyer profession as a proxy for the candidate pool. This proxy is imperfect. First, using the composition of the legal profession as a proxy for the composition of the qualified candidate pool is generous. More women have been attending law school and becoming lawyers over time.<sup>23</sup> The gender balance among young lawyers is more equal than the gender balance of older lawyers, and it is the older, more experienced lawyers who are generally more qualified for prestigious judicial posts.

Second, using lawyers as the proxy for the candidate pool overlooks gender discrimination in informal qualifications.<sup>24</sup> To the extent that women are subject to pressures of the “leaky pipeline,” fewer female lawyers may possess the informal qualifications that increase a judge’s possibility of gaining a seat on the court (Cook 1984; Epstein, Knight and Martin 2003). In addition, gender differences in whether and how candidates perceive themselves as qualified shape decisions to run for office. Studies suggest that women may be less likely to view themselves as qualified (Fox and Lawless 2004; Lawless and Fox 2010), which in turn means that women who do decide to pursue higher office may be *more* qualified than their male peers (Pearson and McGhee 2013; Milyo and Schosberg 2000). Treating all members of the candidate pool as equally qualified discounts gender differences in the accumulation and perception of qualifications.

Despite these limitations, the composition of the legal profession should give us a plausible estimate of the proportion of qualified candidates who are women. If the pattern of women judges replacing women judges systematically suppress diversity on the bench by limiting women to one or a few seats, there would be fewer women state supreme court judges than expected given the candidate pool (all else equal). In contrast, if patterns of women replacing women promotes women judges, we would see more women selected than expected.

---

<sup>22</sup>Based on the author’s research with research assistance from Madeline Brown.

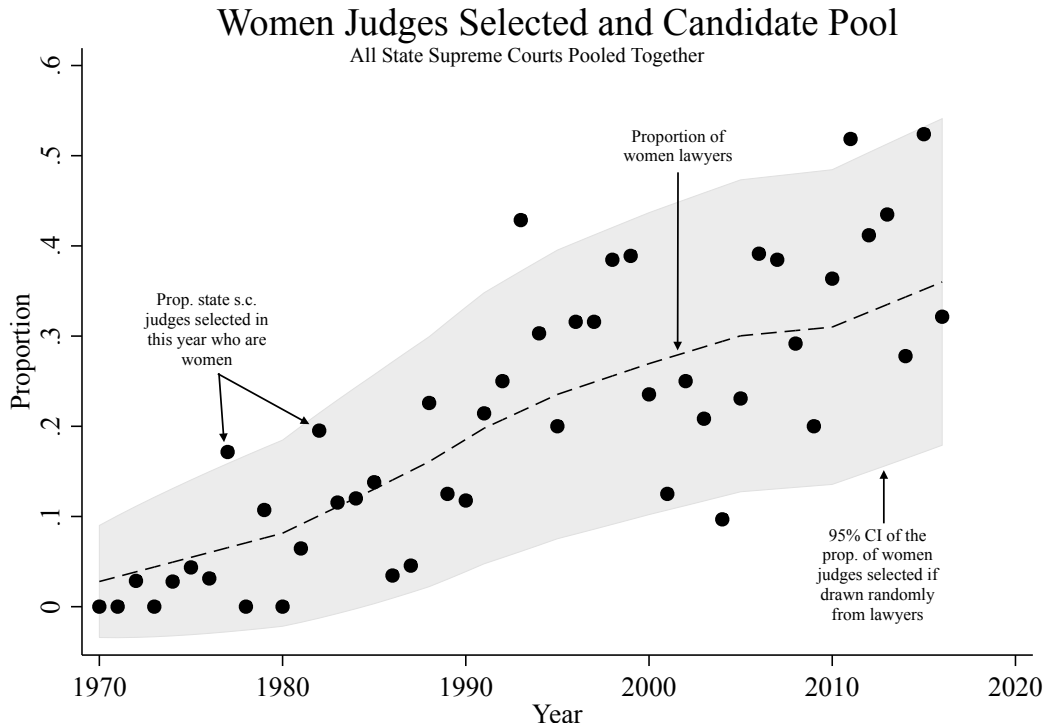
<sup>23</sup>Using a “generous” proxy for the composition of the candidate pool raises the bar for gender diversity, which in turn lowers the bar for observing patterns of exclusion.

<sup>24</sup>See Williams (2006; 2008) and Martin (1981) for discussions of gender variation in formal and informal qualifications for judicial offices.



**Judicial Selection and the Candidate Pool** Figure 3 shows the average proportion of US lawyers who are women (the dashed line).<sup>25</sup> The grey shaded region shows the 95% confidence interval around the composition of the candidate pool.<sup>26</sup> The dots show the actual proportion of vacancies filled by women.

Figure 3



*This plot shows the average percent women in the legal profession (dashed line), the 95% confidence intervals around the proportion of women lawyers (shaded region), and the proportion of women actually selected to state supreme courts (dots) in each year. In all but four years the proportion of women selected to state supreme courts is within the expected range, and three of the four out of range are above the range.*

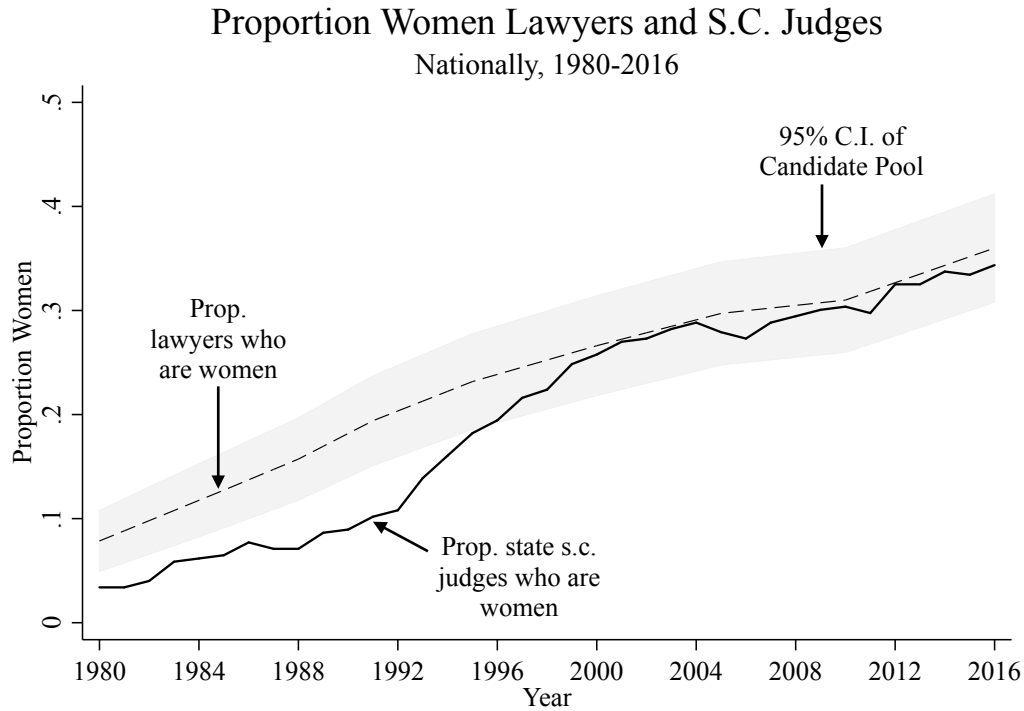
We can see that the proportion of women judges selected to state supreme courts almost always falls within the expected range given the composition of the candidate pool. Of the 46 years in this data, 22 years see a greater proportion of women selected to state supreme courts than the proportion of women lawyers, and 25 years have fewer women judges selected than expected. There

<sup>25</sup>The Lawyer Statistical Report provides state-level data for years 1952, 1963, 1966, 1970, 1980, 1985, 1988, 1991, 1994, 2000, and 2005. These data are aggregated for national comparisons. Data for 2010 and 2016 are only available at the national level. Missing years are estimated with linear interpolation.

<sup>26</sup>That is, if we randomly selected judges from the population of lawyers, 95% of the time the proportion of judges selected that are female would fall within the grey shaded region. The 95% confidence interval is

are only four observations that fall outside the bounds of the confidence intervals, and three of those observations fall *above* the upper bound. This plot shows that the selection of women judges to state supreme courts mirrors a pattern in which judges are randomly selected from the population of attorneys.

Figure 4



*The proportion of women lawyers and the total proportion state supreme court judges on the bench who are women between 1980 and 2016. Data and the number of women selected in each year are more complete in early years than data on the overall composition of state supreme courts. To avoid missing data for the overall composition in early years, I restrict this analysis to 1980*

**Overall Judicial Diversity and the Candidate Pool** Comparing the proportion of judges *selected* to the composition of the qualified candidate pool is more generous than comparing the *overall* proportion of women judges on the bench to the candidate pool: low turnover on courts may depress the overall presence of women on a given bench even if the proportion of vacancies

---

calculated using the following formula for upper and lower bounds:  $p \pm \sqrt{\frac{p*(1-p)}{n}} * 1.96$  where  $p$  is equal to the proportion of women lawyers for a given year and  $n$  is the average number of judges selected each year, 27. I use the average number of vacancies to smooth the confidence intervals for ease of interpretation. A plot that shows the confidence intervals calculated with the specific number of vacancies each year is provided in the appendix. Patterns are the same.

filled by women does reflect the candidate pool. Figure 4 shows the national proportion of lawyers that are women (dashed line) and the national, aggregate proportion of state supreme court judges that are women (solid line). The grey shaded region shows the 95% confidence interval around the proportion women lawyers, which represents variation from the candidate pool that might stem from randomness rather than bias or exclusion. Confirming Cook's (1984) finding that the gender diversity of state supreme court judges lagged behind the candidate pool, the proportion of women judges was below the confidence interval pool prior to 1996. Although the proportion of women state supreme court lawyers has not yet been greater than the proportion women lawyers, the proportion of women judges has been within the expected range for the last 20 years.

Using the gender composition of lawyers over the last 46 years as a proxy for the gender diversity of the candidate pool, women judges have been selected to state supreme court benches as often as expected if gender were not a relevant criterion. While there was a lag in the overall gender diversity of state supreme courts, on average courts have been about as diverse as expected since 1996.

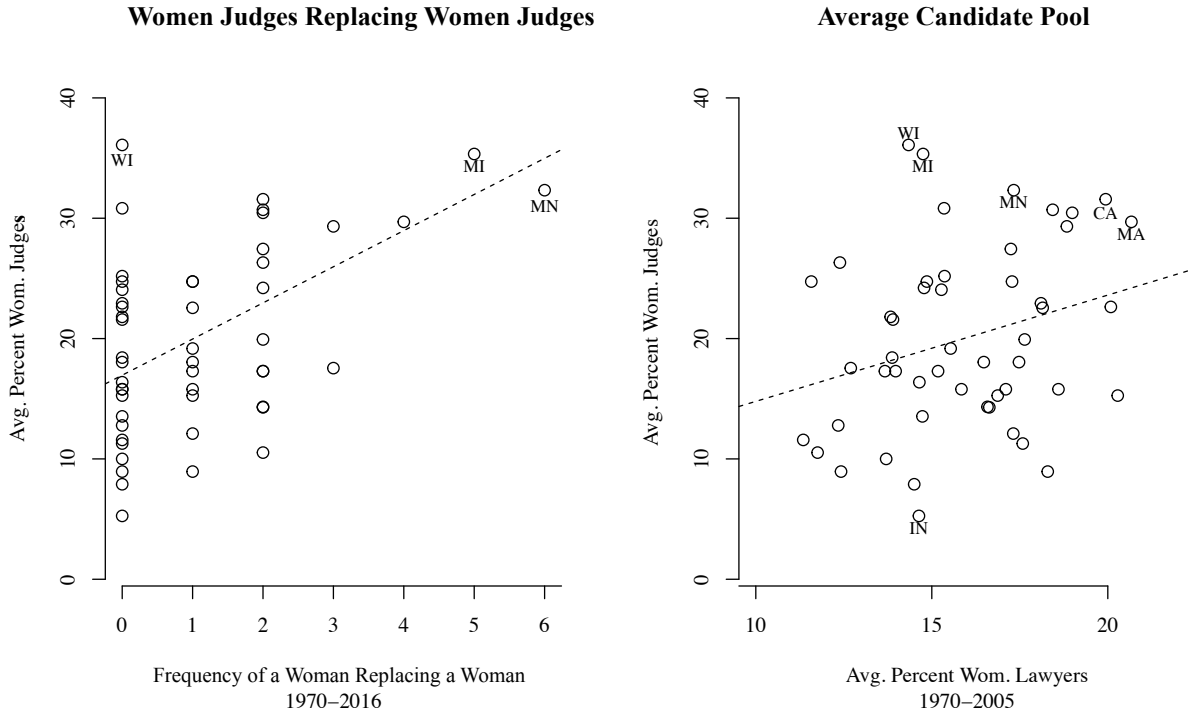
**Gendered Turnover and Diversity** These aggregate patterns of selection and diversity cannot rule out the possibility that gendered patterns of judicial selection limit opportunities for women judges. It could be, for example, that the states that most consistently conform to a pattern of gendered turnover are the states with only one or a few women judges. Empirically, though, that is not the case. The y-axis of the left panel of figure 5 shows the average percentage of women on a state's supreme court between 1970 and 2016. The x-axis shows the number of instances in which a woman judge replaced a woman judge from 1970 to 2016.

The two states with the highest frequency of gendered turnover –Michigan and Minnesota – also are among the most gender diverse. Of course, the relationship between gendered turnover and gender diversity is endogenous: the more women on the court, the more chances there are for women to retire and be replaced by women. Likewise, there are gender diverse courts that do not have patterns of gendered replacement. What is important here is that the patterns of gendered replacement are not limited to courts with minimal diversity, which suggests that this pattern of women replacing women has not been systematically used to limit or exclude women from the

---

<sup>27</sup>State level data on the gender composition of lawyers is only available until 2005.

Figure 5



The left panel shows the relationship between the number of instances in which a woman judge replaced a woman judge by state between 1970 and 2016. The right panel shows the average proportion women judges and the average proportion women lawyers from 1970 to 2005 (2005 was the last year the Lawyer Statistical Report published data about the gender composition of lawyers at the state level).

bench. The right panel of figure 5 shows the average percentage of women judges from 1970 to 2016 and the average proportion of women lawyers from 1970 to 2005.<sup>27</sup> While the relationship is positive, cross-state variation in the gender composition of lawyers does not account for all variation in the gender diversity on the bench. Neither gendered patterns of replacement nor a diverse candidate pool will *necessarily* lead to a diverse supreme court bench, but both factors are positively associated with greater gender diversity.

## Discussion and Conclusions

This project demonstrates that gender is a relevant selection criterion for state supreme courts. Even though there are no formal rules or quotas requiring women to replace women, women judges are more likely to be selected to fill vacancies made by women than vacancies made by men. What

this pattern of gendered replacement means for the overall trajectory of diversification depends on whether patterns of replacement have suppressed or promoted the inclusion of women on the bench. Using the gender composition of lawyers as a proxy for the composition of the qualified candidate pool for state supreme court benches, we see that women are selected to state supreme court benches as often as expected given the candidate pool. These two findings have a few important implications.

First, these findings suggest that the use of gender as a selection criterion has not systematically suppressed the diversification of state supreme court benches at the selection stage. While women are more likely to replace women judges, they do not exclusively replace women.<sup>28</sup> In addition, patterns of gendered judicial replacement are not limited to courts with token levels of diversity, and there is a positive correlation between instances of women replacing women judges and levels of gender diversity on state supreme court benches. Patterns of women replacing women have not resulted in clear patterns of tokenism in which *only* one woman holds a seat on the bench at a time.

While there is no evidence here that gendered patterns of turnover have systematically promoted tokenism at the selection stage, this study does not speak to the experiences of women judges who were selected – or were selected at a specific time – to replace a woman judge. To the extent that people believe women judges are selected in whole or part because of their gender, women judges may be perceived as less qualified than their male peers. Evidence from non-judicial contexts suggests a “discounting principle” in which people perceive beneficiaries of affirmative action policies as less competent (Summers 1991; Heilman, Block and Lucas 1992). Notably, experimental evidence shows that perceptions of incompetence can be offset by unambiguous evidence that beneficiaries of affirmative action policies *are* competent (Heilman, Block and Stathatos 1997). In the judicial context, formal and informal qualification requirements may provide those unambiguous signals of competency; the impressive resumes of women state supreme court justices may preclude potential for observers to “discount” competency. Moreover, Heilman et al. (1998); Evans (2003) find that stigmatization decreases as affirmative action policies become more moderate. The fact that the pattern of women judges replacing women judges is informal and incomplete – not every woman judge who retires is replaced by a woman judge – may serve to alleviate potential discounting.

---

<sup>28</sup>See (Kenney 2012) for an summary of how the replacement of women judges with men decreases overall diversity on the bench in the short term and can be a hurdle to diversification.

Assessing whether gendered patterns of replacement affect perceptions of the competency of women judges or the experiences of women judges will be a fruitful extension of the current project.

Second, gendered patterns of replacement have not resulted in women getting systematically *more* seats than expected given the gender composition of lawyers. Only three times in 30 years has the proportion of women selected been greater than the bounds of the 95% confidence interval, and in those instances the observed proportion of women judges selected was only slightly greater than the expected proportion. If the gender composition of lawyers is an accurate proxy of the candidate pool, then there is no evidence that the use of gender in the selection process systematically favors women judges over men. For advocates of equal gender representation on the bench, however, the selection of women at rates commensurate with the composition of the candidate pool may be viewed as the bare minimum. From this perspective, the selection of women at rates less than 50% is unjust because it reinforces the expectation that women need not be or should not be equally represented in positions of power and provides tacit approval for the current standards for qualification that favor men over women.

Third, this project highlights the importance of continued efforts to diversify the judicial candidate pool and state supreme court benches. While there is no settled standard of gender diversity on state supreme courts, advocates of descriptive representation argue that political bodies ought to mirror the descriptive characteristics of the population. Even though women have been selected as often as expected given the composition of the candidate pool, women do not make up 50% of lawyers or supreme court justices. Women have faced – and continue to face – barriers to the accumulation of formal and informal requirements for office (Cook 1984; Guinier et al. 1994; Drachman 2001; Epstein, Knight and Martin 2003; Redfield 2009; Rikleen 2015). These barriers to accessing qualifications, in turn, suppress the diversity of the eligibility pool. Removing the barriers for women to accumulate the necessary qualifications, or more critically, re-defining what it means for one to be “qualified,” may help lead to more diverse state supreme court benches over time.

Encouragingly, law school graduation rates and the composition of the legal profession are much more diverse now than even 20 years ago. In fact, there were more women law school graduates

---

<sup>29</sup>In addition to continued diversification of the candidate pool, other scholars suggest, for example, that diversifying merit committees (Esterling and Andersen 1999), having women governors (Kenney and Windett 2012), or having more women in legislative office (Hoekstra, Kittilson and Bond 2014) may also lead to increased gender diversity on courts.

than men law school graduates in 2015 and 2016. As these new lawyers mature and acquire informal qualifications, the candidate pool for high office will come to increasingly mirror the gender composition of the electorate. While a more diverse candidate pool may will not necessarily lead to a more diverse judiciary. (Cook 1984), a more diverse candidate pool ought to facilitate efforts to further diversify the judiciary<sup>29</sup>

Finally, although the aggregate patterns of judicial selection conform to expectations, there is cross-state variation in the timing and consistency of diversification and in the use of gendered patterns of replacement.<sup>30</sup> This project has not addressed why some states chose to seek out women to replace women while other states did not. What is it about the social, political, or judicial context in Minnesota and Michigan that accounts for the frequent replacement of women judges with women judges? Of course, having more women on the bench provides more opportunities for gendered patterns of turnover, and Minnesota is unique: its seven judge court had a four woman majority in 1991 (Margolick 1991). To put that in context, Alaska selected its *first* woman judge, Dana Fabe, in 1996 (Alaska Judicial Council 2006), and New Hampshire did not select its first woman judge, Linda Stewart Dalianis, until 2000 (New Hampshire Judicial Branch 2017). Furthermore, within a given state, not all women are replaced by women. Under what conditions are governors and political elites most likely to seek out – either implicitly or explicitly – women judges to replace women judges? Future research ought to address cross-state and over-time variation in the use of gendered patterns of judicial turnover. It is possible that patterns of women replacing women developed out of efforts to diversify the bench but could turn into a ceiling that limits the openness of any seat to a woman judge. Future research should continue to explore whether the judiciary – the branch tasked with ensuring equal justice under the law – is selected through fair, equal, and non-discriminatory practices.

---

<sup>30</sup>See the Appendix for a state-by-state comparison of supreme court diversity and candidate pool diversity.

## References

- Acker, Joan. 1992. "From Sex Roles to Gendered Institutions." *Contemporary Sociology* 21(5):565–569.
- Alaska Judicial Council. 2006. "Dana Fabe."  
**URL:** <https://web.archive.org/web/20060117150532/http://www.ajc.state.ak.us/Judges/FabeDana.htm>
- Alozie, Nicholas O. 1988. "Black Representation on State Judiciaries." *Social Science Quarterly* 69:979–986.
- Alozie, Nicholas O. 1990. "Distribution of Women and Minority Judges: The Effects of Judicial Selection Methods." *Social Science Quarterly* 71(2):315–325.
- Alozie, Nicholas O. 1996. "Selection Methods and the Recruitment of Women to State Courts of Last Resort." *Social Science Quarterly* 77(1):110–126.
- Andrews, Elaine. 2016. "Many Good Reasons for Gov. Walker to Appoint a Woman to the Alaska Supreme Court." *Alaska Dispatch News* .  
**URL:** <https://www.adn.com/commentary/article/many-good-reasons-gov-walker-appoint-woman-alaska-supreme-court/2016/05/03/>
- Atkeson, Lonna Rae. 2003. "Not All Cues Are Created Equal: The Conditional Impact of Female Candidates on Political Engagement." *The Journal of Politics* 65(4):1040–1061.
- Bakalar, Linda. 2016. "Wanting One Woman on Alaska Supreme Court is not 'Reverse Sexism'." *Alaska Dispatch News* .  
**URL:** <https://www.adn.com/commentary/article/wanting-woman-alaska-supreme-court-not-reverse-sexism/2016/05/05/>
- Baldez, Lisa. 2006. "The Pros and Cons of Gender Quota Laws: What Happens When You Kick Men Out and Let Women In?" *Politics and Gender* 2(1):102–109.
- 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. 2001. "Clinton and Diversification of the Federal Judiciary." *Judicature* 84(5):256–261.
- 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.
- Campbell, David E. and Christina Wolbrecht. 2006. "See Jane Run: Women Politicians as Role Models for Adolescents." *The Journal of Politics* 68(2):233–247.
- Carbon, Susan, Pauline Houlden and Larry Berkson. 1982. "Women on the State Bench: Their Characteristics and Attitudes about Judicial Selection." *Judicature* 65(6):294–305.
- Celis, Karen. 2006. "Substantive Representation of Women: The Representation of Women's Interests and the Impact of Descriptive Representation in the Belgian Parliament (1900-1979)." *Journal of Women, Politics and Policy* 28(2):85–114.



- Cook, Beverly B. 1984. "Women Judges: A Preface to Their History." *Golden Gate UL Rev.* 14:573.
- Cook, Beverly Blair. 1978. Women Judges: The End of Tokenism. In *Women in the Courts*, ed. Winifred L. Hepperle and Laura Crites. Williamsburg, VA: National Center for State Courts.
- Cramer Walsh, Katherine. 2002. Female legislators and the womens rights agenda. In *Women Transforming Congress, Congressional Studies Series*. Vol. 4 pp. 370–396.
- Dahlerup, Drude. 2008. "Gender Quotas – Controversial But Trendy." *International Feminist Journal of Politics* 10(3):322–328.
- Dixon, Rosalind. 2009. "Female Justices, Feminism and the Politics of Judicial Appointment: A Re-Examination." *University of Chicago, Public Law and Legal Theory Working Papers* .
- Drachman, Virginia G. 2001. *Sisters in Law: Women Lawyers in Modern American History*. Harvard University Press.
- 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.
- Esterling, Kevin M and Seth S Andersen. 1999. "Diversity and the judicial merit selection process: a statistical report." *Research on Judicial Selection, The American Judicature Society* 2.
- Evans, David C. 2003. "A Comparison of the Other-Directed Stigmatization Produced by Legal and Illegal Forms of Affirmative Action." *Journal of Applied Psychology* 88(1):121.
- 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).
- Fox, Richard L. and Jennifer L. Lawless. 2004. "Entering the Arena? Gender and the Decision to Run for Office." *American Journal of Political Science* 48(2):264–280.
- Gay, Claudine. 2002. "Spirals of Trust? The Effect of Descriptive Representation on the Relationship between Citizens and Their Government." *American Journal of Political Science* 46(4):717–732.
- Goelzhauser, Greg. 2011. "Diversifying State Supreme Courts." *Law and Society Review* 45(3):761–781.
- Greene, Linda S. 2013. "Tokens, Role Models, and Pedagogical Politics: Lamentations of an African American Female Law Professor." *Berkeley Journal of Gender, Law and Justice* 6(1):81–92.
- Guinier, Lani, Michelle Fine, Jane Balin, Ann Bartow and Deborah Lee Stachel. 1994. "Becoming Gentlemen: Women's Experiences at One Ivy League Law School." *University of Pennsylvania Law Review* 143(1):1–110.
- Heilman, Madeline E, Caryn J Block and Jonathan A Lucas. 1992. "Presumed incompetent? Stigmatization and affirmative action efforts." *Journal of Applied Psychology* 77(4):536.
- Heilman, Madeline E, Caryn J Block and Peter Stathatos. 1997. "The affirmative action stigma of incompetence: Effects of performance information ambiguity." *Academy of Management Journal* 40(3):603–625.

- Heilman, Madeline E, William S Battle, Chris E Keller and R Andrew Lee. 1998. "Type of affirmative action policy: A determinant of reactions to sex-based preferential selection?" *Journal of Applied Psychology* 83(2):190.
- Hoekstra, Valerie, Miki Caul Kittilson and Elizabeth Andrews Bond. 2014. Gender, High Courts, and Ideas about Representation in Western Europe. In *Representation: The Case of Women*, ed. Maria C. Escobar-Lemmon and Michelle M. Taylor-Robinson. New York: Oxford University Press book section 3, pp. 103–117.
- Iacus, Stefano M, Gary King and Giuseppe Porro. 2011. "Causal inference without balance checking: Coarsened exact matching." *Political Analysis* .
- Junn, Jane. 1997. *Assimilating or Coloring Participation? Gender, Race, and Democratic Political Participation*. New York: New York University Press book section 22, pp. 387–397.
- Kenney, Sally J. 2008. "Thinking about gender and judging." *International Journal of the Legal Profession* 15(1-2):87–110.
- Kenney, Sally J. 2012. "Choosing Judges: A Bumpy Road to Women's Equality and a Long Way to Go." *Mich. St. L. Rev.* p. 1499.
- Kenney, Sally J. and Jason Windett. 2012. "Diffusion of Innovation or State Political Culture? Explaining the First Woman State Supreme Court Justices." *Working Paper, Cited with Permission* .  
**URL:** <http://2012sppconference.blogs.rice.edu/files/2012/02/KenneyWindettSPPCDraft.pdf>
- Kenney, Sally Jane. 2013. *Gender and justice: Why women in the judiciary really matter*. Routledge.
- Kenworthy, Lane and Melissa Malami. 1999. "Gender Inequality in Political Representation: A Worldwide Comparative Analysis." *Social Forces* 78(1):235–269.
- King, Gary and Langche Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14:131–159.
- Krook, Mona Lena. 2006. "GenderQuotas, Norms, and Politics." *Politics and Gender* 2(1):110–118.
- Lawless, Jennifer L. and Richard L. Fox. 2010. *It Still Takes a Candidate: Why Women Don't Run for Office*. New York: Cambridge University Press.
- Laws, Judith Long. 1975. "The Psychology of Tokenism: An Analysis." *Sex Roles* 1(1):51–67.
- Mansbridge, Jane. 1999. "Should Blacks Represent Blacks and Women Represent Women? A Contingent 'Yes'." *Journal of Politics* 61(3):628.
- Margolick, David. 1991. "Women's Milestone: Majority on Minnesota Court." *The New York Times* .  
**URL:** <http://www.nytimes.com/1991/02/22/news/women-s-milestone-majority-on-minnesota-court.html?pagewanted=all>
- Martin, Elaine. 1981. "Women on the Federal Bench: A Comparative Profile." *Judicature* 65:306–313.

- Martin, Elaine. 1991. *Judicial Gender and Judicial Choices*. Center for the American Woman and Politics: The Impact of Women in Public Office Rutgers.
- Martin, Patricia Yancy, John R. Reynolds and Shelley Keith. 2002. "Gender Bias and Feminist Consciousness among Judges and Attorneys: A Standpoint Theory Analysis." *Signs* 27(3):665–701.
- McDonald, John H. 2014. *Handbook of Biological Statistics*. 3rd ed. Baltimore: Sparky House Publishing.
- Milyo, Jeffrey and Samantha Schosberg. 2000. "Gender bias and selection bias in House elections." *Public Choice* 105(1-2):41–59.
- New Hampshire Judicial Branch. 2017. "Supreme Court - Chief Justice Linda Stewart Dalianis." **URL:** <https://www.courts.state.nh.us/supreme/meetdalian.htm>
- O'Connor, Karen and Jeffrey Segal. 1990. "Justice Sandra Day O'Connor and the Supreme Court's reaction to its first female member." *Women and Politics* 10(2):95–103.
- O'Regan, Valerie R. 2000. *Gender Matters: Female Policymakers' Influence in Industrialized Nations*. Greenwood Publishing Group.
- Paxton, Pamela. 1997. "Women in National Legislatures: A Cross-National Analysis." *Social Science Research* 26(4):442–464.
- Pearson, Kathryn and Eric McGhee. 2013. "What It Takes to Win: Questioning 'Gender Neutral' Outcomes in U.S. House Elections." *Politics and Gender* 9(04):439–462.
- Peresie, Jennifer L. 2005. "Female judges matter: Gender and collegial decisionmaking in the federal appellate courts." *Yale Law Journal* pp. 1759–1790.
- Phillips, Anne. 1995. *The Politics of Presence*. Oxford Political Theory Oxford: Clarendon Press.
- Redfield, Sarah E. 2009. "The Educational Pipeline to Law School: Too Broken and Too Narrow to Provide Diversity." *Pierce Law Review* 8.
- Reingold, Beth. 2000. *Representing Women: Sex, Gender, and Legislative Behavior in Arizona and California*. Chapel Hill: The University of North Carolina Press.
- Reingold, Beth. 2008. *Women as Office Holders: Linking Descriptive and Substantive Representation*. New York: Cambridge University Press book section 9.
- Reingold, Beth and Adrienne R. Smith. 2017. "Who Leads What? Legislative Leadership and Intersections of Gender, Race, and Ethnicity in the American States." *Working Paper* .
- Reingold, Beth and Jessica Harrell. 2010. "The Impact of Descriptive Representation on Women's Political Engagement: Does Party Matter?" *Political Research Quarterly* 63(2):280–294.
- Remiche, Adelaide. 2015. "When Judging is Power." *Journal of Law and Courts* Spring:95–114.
- Rikleen, Luaren Stiller. 2015. "Women Lawyers Continue to Lag Behind Male Colleagues: Report on the Ninth Annual NAWL National Survey on Retention and Promotion of Women in Law Firms." *National Association of Women Lawyers* .

- Saint-Germain. 1989. "Does Their Difference Make a Difference? The Impact of Women on Public Policy in the Arizona Legislature." *Social Science Quarterly* 70(4):956–968.
- Salmond, Rob. 2006. "Proportional Representation and Female Parliamentarians." *Legislative Studies Quarterly* 31(2):175–204.
- Scherer, Nancy and Brett Curry. 2010. "Does Descriptive Race Representation Enhance Institutional Legitimacy? The Case of the U.S. Courts." *The Journal of Politics* 72(1):90–104.
- Schultz, Ulrike and Gisela Shaw. 2013. Introduction: Gender and Judging: Overview and Synthesis. In *Gender and Judging*, ed. Ulrike Schultz and Gisela Shaw. Bloomsbury.
- Solberg, R. L. Spill and K. A. Bratton. 2005. "Diversifying the Federal Bench: Presidential Patterns." *The Justice System Journal* 26(2):119–133.
- Summers, Russel J. 1991. "The Influence of Affirmative Action on Perceptions of a Beneficiary's Qualifications<sup>1</sup>." *Journal of Applied Social Psychology* 21(15):1265–1276.
- 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.
- Wangnerud, Lena. 2009. "Women in Parliaments: Descriptive and Substantive Representation." *Annual Review of Political Science* 12:51–69.
- Williams, Margaret S. 2006. "In a Different Path: The Process of Becoming a Judge for Women and Men." *Judicature* 90(3):104–113.
- Williams, Margaret S. 2008. "Ambition, Gender, and the Judiciary." *Political Research Quarterly* 61(1):68–78.
- Williams, Margaret S. and Frank C. Thames. 2008. "Women's Representation on High Courts in Advanced Industrialized Countries." *Politics and Gender* 4(03).

## Appendix

**Coding Rules, Aggregating Two Years** Because vacancies and selections do not always occur in the same year, I aggregated two years when vacancies and selections did not occur in the same year. Specifically, courts must meet one of three requirements for two years to be aggregated. I implement these rules in order; two years can only be aggregated under rule two, for example, if the two years are not aggregated under rule one. The rules are:

1. If the number of vacancies matched the number of selections in a given year, those selections were paired to those vacancies. Of the 671 state-year-vacancy(s) observations in the unmatched data set, 89 observations were paired under this rule.
2. If there is a vacancy in year  $t$ , no new judge selected in year  $t$ , no vacancy in year  $t + 1$ , but there is a judge selected in year  $t + 1$ , I aggregate the two years so that the judge selected in year  $t + 1$  is counted as the replacement for the judge who retired in year  $t$ . Of the 671 state-year-vacancy(s) observations in the unmatched data set, 497 observations were paired under this rule.
3. In the Bratton data set, judges who take office early in a year but were selected in the previous year are listed as selected in the previous year. Therefore, if a judge retires in year  $t$ , no judge is appointed in year  $t$ , no judge is selected in year  $t + 1$ , no judge vacated in year  $t - 1$ , but a judge was selected in year  $t - 1$ , I count the judge selected in year  $t - 1$  as the replacement to the judge who retired in year  $t$ . Of the 671 state-year-vacancy(s) observations in the unmatched data set, 16 observations were paired under this rule.
4. Finally, I aggregate two years when there is a discrepancy in the number of vacancies and selections in one year, but not over two years. For example, if one judge retires in year  $t$ , no judge is selected in year  $t$ , one judge retires in year  $t + 1$  and two judges are selected in year  $t + 1$ , I aggregate the two years so that the two retiring judges are matched with the two replacement judges. In this case, I treat the aggregated two years as one year with two vacancies. Of the 671 state-year-vacancy(s) observations in the unmatched data set, 69 observations were paired under this rule.

**Disaggregated CMH tests** To determine if the gender of vacating and replacement judges is independent across different conditions, I disaggregate the CMH test across various covariates. Disaggregating across covariates reduces the power of each test. Importantly, all observations are matched on the criteria listed in table 3. The CMH test is flexible to variation across strata, so the aggregated analysis reported in the main text is valid for overall patterns of judicial replacement.

Table 3: C-M-H, disaggregated groups

	Group	# Treated units	$\chi^2$	p
Number Vacancies	Only one Vacancy	51	6.86	0.01
	> 1 Vaancy	7	.15	0.70
# Women on Court	1 Womant	18	3.35	0.07
	> 1 Woman	40	2.82	0.09
	> 2 Women	14	0.32	0.57
Court Size	5 Judge Court	15	.739	0.39
	7 Judge Court	41	4.78	0.03
	9 Judge Court	2	Not enough	data
Selection Method	Popular Election	9	.50	0.48
	App or Merit	48	6.94	0.01
	Legis. Election	1	Not enough	data

**Candidate Pool Confidence Intervals, Actual** Figure 3 plots the candidate pool of lawyers with confidence intervals calculated with the average number of vacancies each year. Figure 7 shows the same plot but with confidence intervals calculated with the actual number of vacancies each year. Patterns are the same.

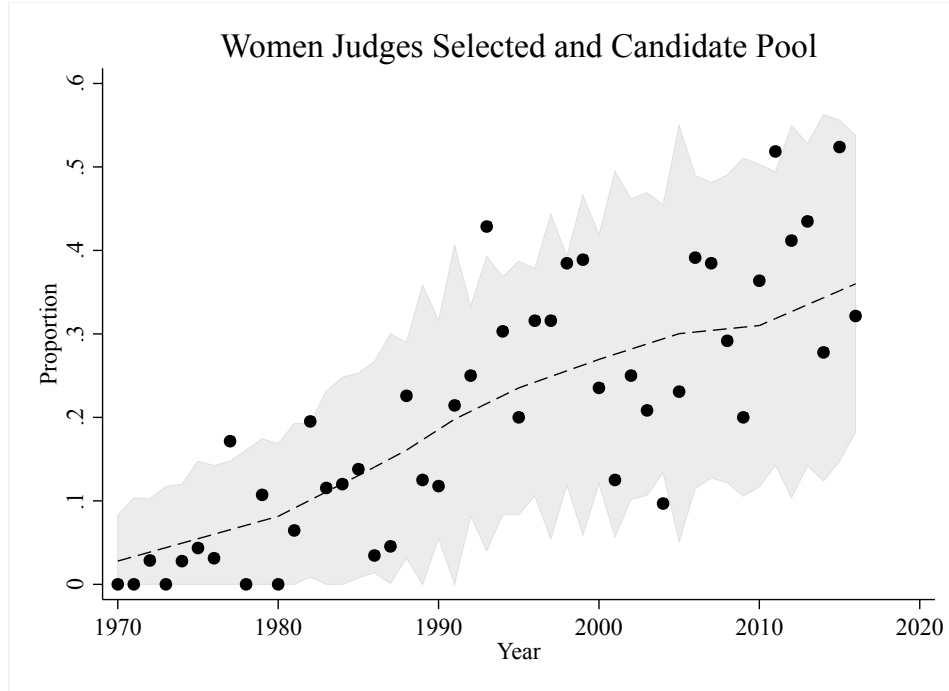


Figure 6: *The dashed line shows the proportion of lawyers who are women over time. The black dots show the proportion of judges selected each year who are women. The shaded region shows the 95% confidence interval for the expected proportion of women selected each year. In this plot the confidence intervals are calculated with the actual number of vacancies each year*

**State-by-state Variation** In addition to obscuring over-time stickiness in diversification, aggregate patterns of selection may obscure variation across state in the selection of women judges. Figure 7 shows how the proportion of women supreme court judges varies by state.

## Proportion Women SC Judges, By State

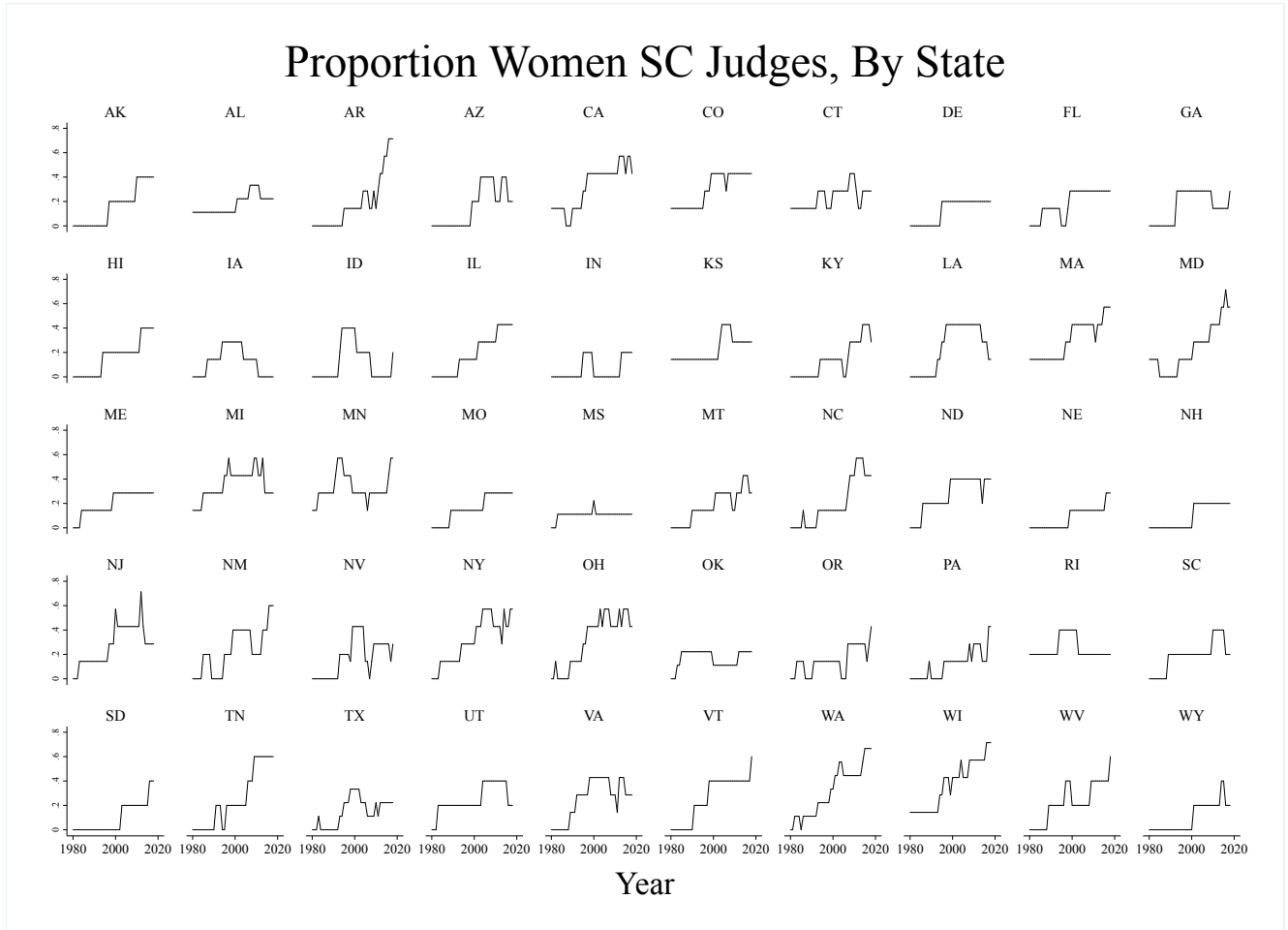


Figure 7: *The proportion of women judges over time and by state. Note: truncating the data at 1980 obscures the trajectories of gender diversity in states where women were selected to the bench earlier: Florence Allen served on the Ohio supreme court from 1923 to 1934; Anne Alpern served on the Pennsylvania Supreme Court in 1961; Lorna Lockwood served on the Arizona state supreme court from 1961 to 1975, and ElsiJane Trimble Roy served on the Arkansas state supreme court from 1975 to 1977.*