

Below the Bar? Racial and Gender Bias in Judicial Nominations*

Maya Sen[†]
msen@fas.harvard.edu

April 30, 2012

Abstract

In this paper, I find that minority and female nominees to federal courts are consistently awarded lower qualification ratings by the nation's largest legal professional organization, the American Bar Association (ABA), than are white and male nominees. This is the case even when comparing nominees with similar pedigrees – including similar education achievements, political ideologies, professional experiences, age, and demographic characteristics. Furthermore, by presenting results showing that the ABA qualification score is unrelated to a judge's ultimate reversal rate, I show that ABA scores are not a particularly good predictor of how a nominee will fare once on the bench. The findings in this paper call into question the ABA's long standing and influential role in the federal nomination process, both in terms of potential latent biases against minority and female candidates and also in terms of its utility in predicting judicial "performance."

*Comments and suggestions welcome. Special thanks to Catalina Santos for exceptional research assistance. This research was supported by the Harvard Center for American Political Studies.

[†]Until June 30, 2012: Department of Government and Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138. After July 1, 2012: Department of Political Science, University of Rochester, Harkness Hall 333, Rochester, NY 14627.

[‡]<http://scholar.harvard.edu/msen>.

1 Introduction

Despite attempts by Presidents and by advocacy groups, federal courts in the United States are still not reflective of the U.S. population. Of the 874 federal judges in service as of 2008, only 24% were women, 10% were African American, and 7% were Hispanic ([Just the Beginning Foundation, 2012](#)). Fewer than 1% were Asian American and, even today, there are no federal judges who self-identify as Native American – surprising given the courts’ involvement in interpreting federal Indian laws. Among legal actors, politicians, and scholars, there is little dispute that the overall population of female and minority judges falls short of being representative of the general population.

Compelling explanations of why descriptive representation in the courts has been so difficult to achieve have eluded social scientists, but a possible contributor is the vetting of presumptive nominees by legal trade organizations such as the American Bar Association (ABA), the nation’s largest and most prestigious lawyers’ association. For example, according to recent accounts, the ABA preliminarily rejected as “not qualified” 14 of Obama’s presumptive judicial nominees. Of these 14 “not qualified” candidates, nine were women and eight were racial or ethnic minorities: all had their candidacies eventually fail ([Savage, 2011](#)). The end result, as some commentators have pointed out, is that the ABA now occupies a quasi-governmental role by systematically “vetoing” certain kinds of candidates. Among liberals and racial and ethnic advocacy groups, the belief is that groups like ABA are biased against minorities and women. Among conservatives, the widespread belief is that the ABA is biased in a liberal direction, a notion that has been confirmed by a handful of empirical papers ([Smelcer, Steigerwalt and Vining Jr, 2011](#); [Lott, 2001](#); [Lindgren, 2001](#)). So strong is this belief that the administration of George W. Bush refused the ABA the long-standing courtesy of “pre-clearing” presumptive candidates before their names were made public and their nominations official.

This paper steps squarely into this debate. Looking at 1,652 judges confirmed since 1960 to the U.S. district courts, I find that black and female judicial nominees are significantly more likely to be awarded lower qualification ratings by the ABA, which increases the likelihood that their nominations will fail. I find that this difference persists after taking into account possible differences in educational and professional backgrounds, age, political ideologies, and years in service. Second, and perhaps more importantly, I also find that the Bar Association scores do little to inform how a nominee will perform once confirmed onto the bench. That is, a judges who are rated as “not qualified” by the ABA are no more likely to have their opinions be overturned once they are invested than are their higher-ranked peers.

Taken together, these findings raise doubts about the process of judicial vetting, and whether an emphasis on prestige credentials (e.g., law school rank) is more important than a close look at political beliefs and partisan affiliations. This finding also calls into question the strong deference paid by political actors to the ABA’s vetting process, and whether individuals who receive a “Not Qualified” rating should have their candidacies summarily withdrawn, as is currently the case ([Savage, 2011](#)). That record numbers of minority and women nominees are currently having judicial candidacies derailed by this vetting process makes this a particularly pressing issue.

This paper proceeds as follows. Section 2 explains how the ABA assesses nominees’ qualifications, focusing specifically on the selection problem that occurs when Presidents decline to move forward with poorly-rated plausible nominees. Section 3 provides an overview of the data used, which are professional and background characteristics of some 1,652 U.S. District Court judges nominated since 1960. I present the key results in Sections 5, 6, and 7, paying particular attention to sensitivity to (1) omitted variables and (2) selection bias. I conclude by discussing the implications of these results.

2 Evaluating Judicial Quality and Possible Bias

Once a judicial vacancy arises, the White House – working closely with the Justice Department and with the senior U.S. Senator from the state with the judicial vacancy – develops a list of presumptive nominees via word of mouth, city bar associations, professional and political organizations, state courts, and area law firms. The short list is then forwarded to the American Bar Association’s Standing Committee on the Federal Judiciary for more vetting. No rule exists mandating that Presidents must present preliminary lists to the ABA for this “pre-clearance”; nonetheless, it has been a long-standing practice followed, with key exceptions, since the Eisenhower administration. Importantly, the list of presumptive nominees is at this point confidential, and the Standing Committee members are prohibited by internal Bar rules from making the names public.

The Standing Committee¹ then begins independently reviewing each presumptive candidate’s record using three criteria: (1) **integrity**, which includes “the prospective nominee’s character and general reputation in the legal community, as well as the prospective nominee’s industry and diligence,” (2) **professional competence**, which “encompasses such qualities as intellectual capacity, judgment, writing and analytical abilities, knowledge of the law, and breadth of professional experience,” and (3) **judicial temperament**, which includes “the prospective nominee’s compassion, decisiveness, open-mindedness, courtesy, patience, freedom from bias, and commitment to equal justice under the law” ([American Bar Association, 2009](#)). The process by which the Standing Committee determines “integrity,” “competence,” and “temperament” is kept strictly confidential, and the Committee does not make any ratings public until the President confirms that the presumptive candidate will be put forward

¹The Committee is composed of 15 individuals from the various federal jurisdictions (known as “circuits”). This includes the Chair of the Committee, two members from the large California-based Ninth Circuit, one member from each of the other 12 circuits. The members appointed by the ABA President for staggered three-year terms and cannot serve more than two terms ([American Bar Association, 2009](#)). Although membership is open to all ABA members, the composition of the Standing Committee has historically been white and male, with its first African American and female members appointed in 1976 and 1977, respectively.

as an official nominee to the Senate Judiciary Committee ([American Bar Association, 2009](#)). Thus, many Presidents have declined to pursue some number of plausible candidacies, possibly based in part on unfavorable (yet not publicly disclosed) preliminary ABA ratings. This practice could introduce substantial bias into the analysis and raises the possibility of implicit bias by the ABA when none in fact exists. I discuss this below.

The opacity of the ratings process has led to assertions that certain candidates are systematically disadvantaged. In this regard, the strongest critique has been that the ABA is biased left-ward and that ideologically conservative candidates and/or candidates nominated by Republican presidents are more likely to receive lower ABA ratings. Examining data from two administrations, for example, [Lindgren \(2001\)](#) finds that confirmed Bill Clinton appeals court appointees with no judicial experience had “9.7 times as high odds of getting the highest ABA rating” as similar George H.W. Bush appointees, controlling via logit regression for key differences. Although [Lindgren \(2001\)](#) finds no differences between nominees *with* judicial experience he does find differences in the criteria that are predictive of high ABA marks under the Clinton and Bush I regimes. (These findings were later critiqued by [Saks and Vidmar \(2001\)](#) on the grounds that the analysis did not include presumptive nominees, as well as District Court nominees, and could therefore be biased.) Similar results are obtained by [Lott \(2001\)](#), who does collect data from a handful of presumptive appeals court nominees whose names were not put forward as actual candidates. More recently, scholarly evidence in favor of a partisan bias has been put forth by [Smelcer, Steigerwalt and Vining Jr \(2011\)](#), which uses genetic matching to find a bias against Republican Court of Appeal nominees. They find, however, no evidence associated with either race (non-white status) or gender.

That the ABA could be partisan or biased against ideological conservatives has had substantial political ramifications. The Federalist Society, a right-leaning legal organization, publishes a newsletter entitled “ABA Watch” in which it closely monitors potentially bi-

ased treatment of conservative candidates by the Bar,² and numerous commentators and influential bloggers have also weighed in to provide anecdotal evidence on this issue (e.g., [Whelan \(2010\)](#); [Mirengoff \(2010\)](#); [Lott \(2006\)](#)). Conservative ire at the ABA crested in 2001 when George W. Bush’s Attorney General, Alberto Gonzales, notified ABA President Martha Bennett that the White House would no longer allow the ABA to preview and vet the confidential short-list of presumptive candidates before the nominations became official ([Gonzales, 2001](#)). Thus, during the entirety of George W. Bush’s administration, candidates were nominated regardless of the ABA’s rating, and the ABA only had access to the same list of *actual* nominees that Congress, the media, and the public did. (Following the inauguration of Barack Obama in 2009, the custom of allowing the ABA to review the short list of presumptive candidates privately, before the names were made public, was resumed.) Because the Bush II era essentially circumvents the selection bias problem that plagues other administrations, I leverage these 261 nominees in my analysis below.

Comparatively less attention has been paid to the relationship between American Bar Association ratings and race and/or gender. [Lott \(2001\)](#) notes in passing that African American appeals court nominees – in particular African American *Republicans* – are most likely to get lower ratings, although these findings do not go to the core of his results; [Smelcer, Steigerwalt and Vining Jr \(2011\)](#), on the other hand (and to their surprise), find no statistically significant relationship between race or gender and ABA qualification ratings. Anecdotally, however, the belief has increasingly been that the Bar is tilted against some of these candidates, perhaps owing to women and minorities having less “courtroom” experience and more government and/or academic experience ([Savage, 2011](#)). Obama Administration officials have, for example, have been confidentially informed that the ABA has so far “opposed 14 of the roughly 185 potential nominees the administration asked it to evaluate.” Of these “nine are women – five of whom are white, two black, and two Hispanic.

²<http://www.fed-soc.org/publications/page/aba-watch>.

Of the five men, one is white, two are black, and two are Hispanic” ([Savage, 2011](#)).

This perceived negative treatment of minority candidates has, furthermore, led to tensions between the ABA and Democrats and liberal advocacy groups. (To this extent, the ABA has found itself in opposition with an unlikely combination of conservatives and liberals.) The Obama administration has declined to pursue the candidacies of some of the presumptive nominees preliminarily deemed by the ABA as being “Not Qualified,” which has led to concerns about the success of its diversity initiatives ([Savage, 2011](#)). Senator Harry Reid claimed that the ABA needed to “get a new life” following its awarding of a low rating to Obama nominee Gloria Navarro ([Tetreault, 2010](#)), who was later confirmed by the Senate by a vote of 98-0. And speaking about Latina nominees specifically, Robert Raben, a member of the left-leaning American Constitution Society, wrote in a recent op-ed that

I have not seen a single Latina nominee who wasn’t either hit or slammed by some establishment group – a bar association, a leader of a not for profit, a bar leader, a judicial committee – as being ‘intemperate’; lacking “seasoning”; “inexperienced”, “not that bright”, etc etc....There’s a possibility that the entire cohort of Latina lawyers who want to be federal or state judges just don’t deserve it yet, but I’m not buying it. I think there’s something else going on, and I think that unearthing what may be going on within the ABA’s cloistered process may help us get to the bottom of this ([Raben, R., 2011](#)).

3 Data

The sample of interest is 1,652 U.S. District Court judges nominated between 1960 and early 2012.³ (I start the clock at 1960 because the first African American district judge was

³I also plan on including in the analysis approximately 200 withdrawn nominations, for which I only have preliminary data at the time of writing. All data and accompanying statistical code will be posted to a replication archive at the conclusion of this project.

confirmed in 1961, and there is no support for cross-race comparisons, and very little support for cross-gender comparisons, before 1960.) I choose the district courts as opposed to higher levels of the federal judiciary due to its size. Compared to the nine Justices serving on the Supreme Court, and to the approximately 180 judges serving on the U.S. Appeals courts (the middle level of the federal courts), approximately 700 judges serve at any given point on the U.S. District courts. This wealth of data allows us to more systematically analyze discrepancies in confirmation outcomes on the basis of sex, gender, or political affiliation. This also makes this study distinct from earlier studies – e.g., [Smelcer, Steigerwalt and Vining Jr \(2011\)](#); [Lott \(2001\)](#); [Lindgren \(2001\)](#) – which focus on appeals court judges.

In addition, the U.S. District Courts provide a good basis for understanding whether external qualification ratings predict judicial “performance.” Of the nearly 300,000 cases per year filed in district courts, around 70,000 are appealed to the U.S. Courts of Appeals, which then reverse or uphold the lower-court judges’ rulings. These rulings provide a convenient population to analyze separately: we can determine whether a lower court judge’s ABA rating will be predictive of his or her reversal rate. This contrasts with the appeals courts, from which only approximately 70 cases per year are appealed to the U.S. Supreme Court.

For each of the 1,652 district court judges, I collected his or her ABA qualification rating using biographical data provided by the Federal Judicial Center.⁴ The ABA currently awards three possible ratings: (1) **Well Qualified**, for which “the prospective nominee must be at the top of the legal profession in his or her legal community; have outstanding legal ability, breadth of experience, and the highest reputation for integrity; and demonstrate the capacity for sound judicial temperament,” (2) **Qualified**, in which the nominee “satisfies the Committee’s very high standards with respect to integrity, professional competence and judicial temperament, and that the Committee believes that the prospective nominee is qualified to perform satisfactorily all of the duties and responsibilities required of a federal judge,” and

⁴<http://www.fjc.gov>.

	Not Qualified	Qualified	Well Qualified	Exceptionally Well Qualified	<i>N</i>
All	0.01	0.43	0.54	0.02	1652
Whites	0.01	0.41	0.56	0.03	1388
Blacks	0.01	0.57	0.41	0.00	147
Hispanics	0.02	0.56	0.41	0.01	95
Women	0.00	0.47	0.52	0.00	279
Men	0.01	0.42	0.55	0.03	1373
Democrats	0.01	0.42	0.54	0.02	726
Republicans	0.00	0.43	0.54	0.02	926

Table 1: Distribution of ABA Qualification Ratings for U.S. District Court Judges confirmed after 1960.

(3) **Not Qualified**, where the ABA has “determined that the prospective nominee does not meet the Committee’s standards with respect to one or more of its evaluation criteria – integrity, professional competence or judicial temperament” ([American Bar Association, 2009](#)). Two other categories have been discontinued: (4) **Exceptionally Well Qualified**, discontinued in 1989, and (5) **Not Qualified by Reason of Age**, discontinued in 1980. Only three confirmed judges ever received the “Not Qualified by Reason of Age” rating, which was automatically awarded to individuals over the age of 60 at the time of nomination. Because so few nominees received this rating, and because this rating was deterministic, I drop it from the analysis.

A demographic breakdown of scores by race, gender, and party affiliation (by party of the appointing President) is provided by Table 1. Very few judges – only about 3% – were ever awarded the two most extreme categories, “Exceptionally Well Qualified” and “Not Qualified.” About 43% have been awarded the second lowest category, “Qualified,” with the majority of judges, 54%, being awarded the second highest category “Well Qualified.” (The same is, however, not true for minority judges, more of whom were awarded the lower “Qualified” category: 57% of African Americans and 56% of Hispanics received this category.) Because so few nominees were ever awarded the two most extreme categories, and because the highest category (“Extremely Well Qualified”) was abolished in 1989, I move

forward by dichotomizing the qualification scheme into two categories: (1) those who received one of the highest two categories versus (2) those who received one of the lowest two categories. Dichotomizing the ABA scores in this way is routinely done in this literature, and never changes the inferences about the middle two categories – “Well Qualified” and “Qualified.”

In addition to recording a judge’s ABA rating, the data from the Federal Judicial Center include demographic characteristics such as age, place of birth (or death, if applicable), law school attended, past judicial experience, and a brief blurb describing the judge’s previous professional experience. Because previous professional experience speaks directly to the ABA’s criteria of “professional competence,” I used automated content analysis to code these excerpts to indicate whether each judge had (1) legal clerkship experience,⁵ (2) had worked in private practice, was (3) a full-time law professor or dean,⁶ (4) worked as Congressional counsel or as (5) an attorney with the Department of Justice, and whether the judge was ever (6) a U.S. Attorney or (7) an Assistant United States Attorney. I also coded whether the judge had worked in a judicial capacity before – for example, as a federal magistrate, bankruptcy, or territorial judge, or as a state judge (both state lower court and state supreme court judge). The breakdown by race, gender, and party affiliation is reported in Table 2.

The Federal Judicial Center also includes each judge’s gender and race or ethnicity. The racial/ethnic categorizations used by the Judicial Center are mutually exclusive, relying on self-identification, and include white, African American, Hispanic, Asian American, and Native American. (The Judicial Center therefore treats “Hispanic” as a distinct racial categorization.) Also reported is the law school and undergraduate institution each judge attended.

⁵I define this as whether the judge clerked for an individual judge, as opposed to serving as a court clerk, clerk of the court, or court staff attorney, occupations that sometimes were sometimes designated by the Federal Judicial Center as “law clerk.”

⁶Here, I exclude individuals who worked as adjunct or visiting professors, lecturers, or clinical instructors.

	Whites	Blacks	Hispanics	Women	Men	Democrats	Republicans
Ave Age at Investiture	50.43	48.55	47.66	47.93	50.50	50.59	49.65
Female	0.15	0.27	0.28	-	-	0.23	0.12
Appointed by Democrat	0.40	0.71	0.49	0.61	0.40	-	-
Top 14 Law School	0.30	0.28	0.24	0.29	0.30	0.32	0.29
Private Law School	0.51	0.67	0.44	0.59	0.51	0.54	0.51
Law Clerk	0.22	0.14	0.12	0.35	0.18	0.23	0.20
Law Professor	0.05	0.12	0.06	0.07	0.06	0.07	0.05
Private Practice	0.94	0.76	0.84	0.82	0.94	0.91	0.92
US Attorney	0.09	0.03	0.05	0.06	0.09	0.06	0.10
Assistant US Attorney	0.19	0.29	0.21	0.29	0.18	0.19	0.21
Justice Dept Lawyer	0.05	0.07	0.04	0.06	0.05	0.05	0.05
Public Defender	0.03	0.10	0.14	0.06	0.04	0.07	0.02
US Magistrate Judge	0.08	0.10	0.15	0.20	0.07	0.09	0.08
US Bankruptcy Judge	0.01	0.04	-	0.03	0.01	0.01	0.01
State Judge	0.38	0.55	0.50	0.45	0.40	0.42	0.40
<i>N</i>	1388	147	96	279	1373	726	926

Table 2: Demographics of U.S. District Court Judges confirmed after 1960.

4 Methodology

Because minority and female nominees on average have differences in terms of their legal training, professional backgrounds, and judicial experience (Table 2), simple comparisons may mask substantial differences in these populations. To account for possible differences, I rely on matching (Ho et al., 2007). Matching allows the comparison of nominees who are identical across key characteristics. Thus, a female nominee who graduated from a Top 14 law school and who previously served as federal magistrate judge will be compared with a comparable male nominee who also graduated from a Top 14 school and who also worked as a federal magistrate judge.

This approach offers several advantages. First, matching is an effective pre-processing step that reduces dependence on statistical modeling assumptions (Ho et al., 2007). Second, and relatedly, matching effectively tests all possible ways that variables could interact with each other. We may, for example, think that the ABA might treat male and female judges

differently, but only among individuals attending lower-ranked law schools. By pruning the data, matching resolves this problem and isolates the effect of a nominee being female or African American, regardless of the possible ways that other variables may be affecting one another. To implement the matching, I use coarsened exact matching (Iacus, King and Porro, 2011, 2009), which allows exact matching on key variables and coarsening and then matching approximately on the few variables that are continuous (discussed below). Coarsened exact matching has the advantage of allowing for this approximation to be as close as needed to remove biases. I also have the advantage of matching exactly – the best form of matching – on a large portion of the variables. Once nominees were matched, I took the difference in means in their ABA ratings.

As discussed below, however, we may be interested in estimating the differences in ratings assigned over not just the subset of the population for which there is overlap in professional characteristics (e.g., the matched sample), but also over the full population of interest (e.g., all nominees). We may also be interested in how the ABA ratings differ across certain population subsets – including across different party affiliations or across different geographic jurisdictions; these may all have implications for the causal mechanism(s) behind the results. Lastly, as discussed above, we may be interested in looking at whether (and to what extent) ABA qualification ratings could be useful predictors of a judge’s performance once confirmed on the bench. Thus, I at times fit logit models, in most instances controlling for the same variables used in the matching. In addition, because the coefficients obtained using a logit link function can be difficult to interpret, I present predicted probabilities throughout. The substantive results of these models reinforce the results from the matching.

At all times I match on, or control for, key personal characteristics of the U.S. District Court nominees, including whether the nominee (1) was a former law clerk, (2) had ever served as a United States attorney or as an Assistant United States attorney, (4) had worked in the Solicitor General’s Office (as a Deputy or Assistant Solicitor General), (5) had ever

served as a state judge (either as a state supreme court or state lower court judge), (6) had ever been a former federal judge (e.g., magistrate, territorial, or bankruptcy judge), (7) had worked as a full-time law professor or law school dean, (8) had experience as an attorney in private practice, or (8) had ever been a public defender.⁷ I also match the judges' law school 2001 U.S. News & World Report rankings, dividing them into six categories: (1) elite law schools in the "Top 14," (2) other law schools in the Top 25, (3) law schools ranked between 26-50, (4) law schools ranked between 51-76, (5) law schools ranked 76-100, and (6) law schools ranked outside of the top 100. (These are admittedly a somewhat rough measure for judges attending law school in the 1960s and 70s.) I also include a dummy variable for whether the law school was public or private and include in the analysis the nominee's age, coarsening to create four age cohorts: (1) 30-40, (2) 40-45, (3) 46-55, and (4) 55+.

In terms of political ideology, I include two key variables (where appropriate). First, I match on, or include dummy variables for, the President who nominated the judge. This also has the effect of conditioning on administration idiosyncrasies and possible fluctuations due to external historical or social trends. Second, for those nominees who went on to be confirmed and invested (i.e., those who went on to become district court judges), I use the judges' judicial common score (Boyd, 2011; Epstein et al., 2007; Giles, Hettinger and Peppers, 2001; Poole, 1998), which relies on the party of the appointing President or, if the party of the appointing President coincides with that of the senior senator of the nominee's

⁷Making causal claims pertaining to immutable characteristics is particularly thorny because the "treatment" (e.g., race or ethnicity, gender) is assigned at birth, rendering (1) experimental analogies ill defined and (2) nearly all control variables post-treatment (Greiner and Rubin, 2010; Sen and Wasow, 2011). Here, I conceptualize the treatment not as being the nominee's race or gender, but the *exposure* of American Bar Association's Standing Committee on the Federal Judiciary to the nominee's immutable gender, race, or ethnicity (Sen and Wasow, 2011). Thus, a well defined experiment would be taking identical nominees (with identical profiles) and randomly assigning the "race" associated with the nominees – similar to what is done in audit studies in public health, housing, and labor economics. Conceptualizing the observational study in this way highlights that the moment of "treatment" happens when the nomination packet is assembled and initially presented to the Bar Association. Lastly, it is also important to note that conceptualizing the "treatment" as being assigned at birth does not actually affect the core findings: black, female, and Hispanic nominees still receive lower ABA scores even when no statistical controls are included (Table 1).

	All	Whites	Blacks	Women	Men	Democrat	Republican
Female	0.17	0.00	0.00	1.00	0.00	0.03	0.03
Democrat Appointment	0.44	0.61	0.61	0.72	0.72	1.00	0.00
Top 14 Law School	0.30	0.33	0.33	0.69	0.69	0.36	0.36
Law Professor	0.06	0.00	0.00	0.03	0.03	0.01	0.01
Private Practice	0.92	1.00	1.00	0.97	0.97	0.99	0.99
Assistant US Attorney	0.20	0.11	0.11	0.19	0.19	0.06	0.06
Justice Dept Lawyer	0.05	0.07	0.00	0.06	0.07	0.04	0.03
Law Clerk	0.21	0.17	0.17	0.31	0.31	0.11	0.11
US Magistrate Judge	0.09	0.00	0.00	0.03	0.03	0.01	0.01
US Bankruptcy Judge	0.01	0.00	0.00	0.00	0.00	0.00	0.00
State Judge	0.41	0.67	0.67	0.34	0.34	0.40	0.40
Ave Commission Year	1988.37	1988.95	1988.56	1997.31	1997.21	1983.62	1985.89
<i>N</i>	1652	24	18	32	37	362	499

Table 3: Pre-matching (for all judges) and post-matching characteristics, for (1) blacks compared to whites, (2) women compared to men, and (3) Democrats compared to Republicans.

state, the common score of the senior senator. (For those nominees whose candidacies were withdrawn, I intend to create my own common score by also using either the ideological common score of either the appointing President or the senior senator of the state.)

A summary of lower-court judge characteristics post-matching is given by Table 3. This matched sample of judges is, as expected, slightly different than the original pre-matched sample (the first column of Table 3 and as well as Table 2) but certainly not fundamentally atypical. Very few of the matched judges had experience working as magistrates or bankruptcy judges, as law professors, or as Assistant U.S. Attorneys, a testament to the small number of such individuals in the population of judges at large. In addition, the matched sample has (in most instances) a greater proportion of individuals who attended a Top 14 law school, whose careers were spent in private practice, and who were nominated by Democrats. Lastly, the average commission year fluctuates somewhat from the overall sample, reflective of the fact that certain candidates (e.g., women) are nominated more frequently in later administrations.

5 Predictors of ABA Ratings

I begin by showing how various judicial characteristics are predictive of the ABA ratings awarded. Here and in subsequent analyses the outcome variable is whether the nominee was highly rated by the ABA, receiving either a “Exceptionally Well Qualified” or “Well Qualified” rating. Thus, Figure 1 shows the relationship between key professional characteristics and whether a nominee earned one of the higher ABA ratings versus one of the lower ones. (Inferences about the two middle categories do not change using an ordered logit specification.) I also include dummy variables for race or ethnicity (with whites comprising the baseline group), gender, and appointment by a Republican, which are the variables of interest in the analysis. Other controls include the age of the nominee and the rank cohort of the nominee’s law school. (Dummy variables for the identity of the appointing President do not change the results; these are reported in Table 11 in the Appendix.)

As the results from Figure 1 suggest, certain traits are positively linked with earning a high ABA score. For example, individuals who have previous judicial experience (e.g., previous experience as a state judge, a U.S. Bankruptcy Judge, or a U.S. Magistrate judge) are more likely to receive the higher two ratings. Other characteristics that are linked with higher scores include whether the judge attended a Top 25 Law School, spent time in private practice, or served as a U.S. Attorney. Age (here measured in years at time of commission) is also positively associated with receiving a higher ABA score. Two other characteristics – whether the judge was a former law clerk and whether the judge attended a private law school – are also positively linked with higher ABA ratings, but fall just shy of statistical significance. Thus, we can identify that prestige (e.g., high law school rank) matters, as does practical experience – including private practice experience and judicial experience.

Three traits are negatively linked with receiving low ABA ratings. These include a judge being (1) female, (2) African American, or (3) Hispanic. Effects for all three are statistically

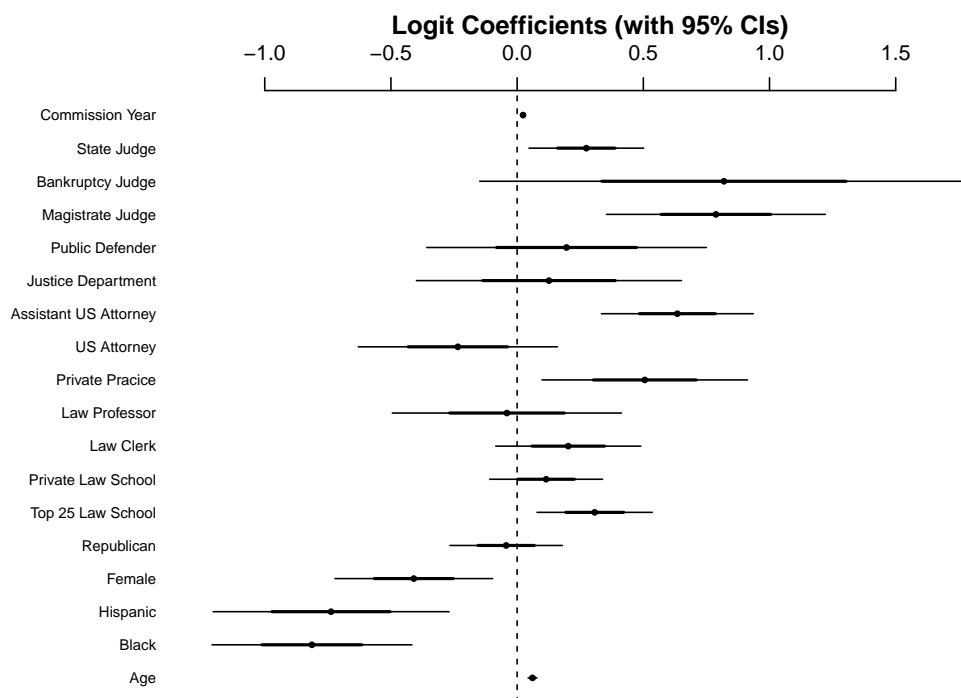


Figure 1: Logit regression results. Outcome variable is Exceptionally/Well Qualified rating (versus Not Qualified/Qualified rating) from the ABA. Solid dots represent point estimates; lines present one- and two-standard deviation intervals.

significant. (There were insufficient numbers of Asian American or American Indian judges to make meaningful inferences about these groups.) A fourth variable of interest, a judge being Republican, does not seem to have much of a relationship with the rating awarded, and it is not statistically significant under any model specification. Thus, at a preliminary level, we see some evidence that women and racial ethnic minorities receive lower ABA ratings than men and white nominees, even after controlling for key judicial characteristics, and no support for ideological differences based on the party of the appointing President.

6 ABA Ratings, Racial Minorities, and Women

The results presented in Figure 1 suggest that racial/ethnic minorities and women are receiving lower scores, even after controlling for key judicial characteristics. However, it could be the case that these results are model dependent, possibly obscuring the true role played by race/ethnicity and gender. Relatedly, it is also possible that the model is making predictions outside of the support of the data – e.g., in instances where there are no substantially comparable whites and blacks, or men and women. Thus, I now turn to matching to more closely investigate the potential role played by race and gender.

Unfortunately, because there are so few Hispanics, and nearly no Asian Americans (and no Native Americans), I focus the race/ethnicity part of the analysis on African Americans, here compared to whites. In each instance, unless noted otherwise, I first match on the relevant personal and professional characteristics, including (1) judge gender (or race in the case of women, discussed below), (2) the identity of the appointing president, (3) age (using four age cohorts), (4) state judge, (5) U.S. Attorney, (6) Assistant U.S. Attorney, (7) Solicitor General Assistant or Deputy, (8) Federal Magistrate or Bankruptcy Judge, (9) law professor, (10) private practice experience, (11) public defender experience, (12) law clerk experience, (13) law school rank, and (14) ideology (as measured by the nominee’s judicial common score). Next I calculate the difference in means in the two populations (black and white nominees) in terms of the ABA rating awarded.

Results from after matching on these key characteristics are presented in Table 4. Looking at African Americans, for example, an estimate of -34% indicates that black judicial nominees are on average 34% *less* likely to receive a high rating from the ABA than are professionally similar whites nominated by the same Presidents, a difference that is also statistically significant at the 5% level with 95% confidence intervals of -55% to -8%. (Conversely, the same analysis also results in African Americans being 34% *more* likely than

	Prob Change in Receiving High Rating	95% CI
African Americans	-0.34	(-0.55, -0.082)
Women	-0.24	(-0.43, -0.036)

Table 4: Change in probability, after matching, of receiving one of the highest two ratings from the ABA.

similarly situated whites to receive the lowest two ratings, and that difference is also statistically significant.) Different coarsening and including other professional factors into the analysis never change the direction or even rough magnitude of the results.

The results attenuate slightly for female nominees. For women, I match them to men across the same characteristics as before; the one exception is that instead of matching on the nominee’s gender, I match on the nominee’s race or ethnicity (so as to hold that constant). The results, presented in the second row of Table 4 demonstrate that women are, on average, 24% less likely than similarly situated men to receive a high rating from the ABA. Women are also more likely to receive the lower two ratings awarded by the American Bar Association (“Not Qualified” and “Qualified”). Both findings are statistically significant.

Sensitivity to Omitted Variables

Although I match on, or otherwise take into account, a substantial number of factors that could possibly influence the scores awarded, it is clearly possible that (1) we do not have access to the full breath of information available to the ABA’s Standing Committee on the Federal Judiciary or that (2) some of the information used by the ABA is inherently qualitative in nature and not included in the Federal Judicial Center’s amalgam of data.

To gain some traction over the possibility that unobserved covariates are driving the results presented in Table 4, I use a method of sensitivity analysis described by Rosenbaum (2002), implemented in R using the `rbounds` package developed by Keele (2010). This sensitivity analysis works roughly by hypothetically “increasing” the level of unobserved

	Post-Matching Coefficient	p -value	Γ Statistic
African Americans	-1.63	0.0022	1.70
Women	-1.10	0.006	1.25

Table 5: Original post-matching logit coefficient estimate, exact p -value under no confounders, and Rosenbaum sensitivity analysis Gamma value.

covariate(s) in the “treated” population (e.g., racial and ethnic minorities, women) until the results are no longer significant. Thus, the sensitivity analysis gives us an estimate of the size of the bias (denoted as Γ) that must be present in these populations in order for the results to be called into question. For example, a result of $\Gamma = 1.2$ for African American nominees means that there must be 20% more of some unobserved trait among the African American nominees for the results to lose significance. Although there is no firm agreement in the literature about the minimum Γ value for observational studies, anything above $\Gamma = 1.5$ appears to indicate substantial insensitivity to unobserved confounders.

The results are presented in Table 5 and demonstrate that (by observational standards) the results are actually fairly insensitive. In order to yield the results insignificant, some trait would have to be present in the African American judge population approximately 1.70 times as often as in the white population. Given that the analysis already controls for clerkship experience, professional experience, quality of legal education, and previous judicial experience, this seems unlikely. It could be the case that, for example, African American judges are 70% less likely than white judges to have been on their school’s law review or to have graduated as members of the Order of the Coif, a law school honors society. Controlling for the law school’s ranking and for subsequent judicial experience (for which such metrics might be predictive), again, makes this somewhat unlikely. For women, the results are more sensitive, a result consistent with the smaller treatment effect for this group (Table 4). In order to yield the results insignificant, women nominees would have to have some treatment approximately 1.25 times as often as male nominees.

Sensitivity to Selection Bias

As noted, the ABA makes public its qualification ratings only for those individuals who were eventually nominated by the White House and whose candidacies advanced to the Senate Judiciary Committee; that is, ABA qualification scores are available only for *actual* nominees, not *presumptive* nominees ([American Bar Association, 2009](#)). In addition, the Federal Judicial Center collects the ABA scores and previous professional and judicial experience of those nominees who were actually nominated by the White House, confirmed by the Senate, and invested as U.S. District Court judges. Thus, both the Federal Judicial Center data (and additional data collected by [Zuk, Barrow and Gryski \(2009\)](#)) systematically exclude ABA ratings of (1) individuals whose candidacies were dropped during the ABA’s “pre-clearance” stage, and (2) individuals who were actually nominated by the White House but whose nomination eventually failed or was withdrawn.

Although not publicized, anecdotal evidence suggests that the actual number of failed presumptive nominees appears to be quite small, somewhere around 3-5 per four-year term.⁸ A significant concern is, however, that not including these individuals in the analysis could bias the results. For example, it could be the case that Presidents starting with Jimmy Carter were eager to appoint minority judges, perhaps in order to increase more rapidly the proportion of black and women judges on the courts. Under such a scenario, it is quite possible that Presidents who had their “short lists” vetted by the ABA would move forward by officially nominating “Not Qualified” minority or female candidates to the full Senate, while declining to move forward the nominations of “Not Qualified” white or male candidates. The same is true for individuals who were actually nominated by the White House but were rejected by the Senate or withdrew their nominations. In that context,

⁸According to [Lott \(2001\)](#), “three potential nominees were said to have been advised that they would get a ‘not qualified’ rating during Bush I and nine potential nominees fell into this category for Reagan”; Bush II did not submit names for ABA “pre-clearance,” while Obama, an exception, has had about 14 nominees whose names have not moved forward due to receiving a poor ABA mark ([Savage, 2011](#)). The identities of these failed presumptive nominees is strictly confidential.

President Name	Whites	African Americans	Hispanics	Women	<i>N</i>
Barack Obama*	0.72	0.17	0.11	0.47	108
George W. Bush	0.82	0.07	0.11	0.21	261
William J. Clinton	0.76	0.18	0.06	0.29	305
George H.W. Bush	0.89	0.07	0.04	0.2	148
Ronald Reagan	0.93	0.02	0.05	0.08	290
Jimmy Carter	0.78	0.15	0.07	0.15	195
Gerald Ford	0.91	0.06	0.02	0.02	49
Richard M. Nixon	0.96	0.03	0.01	0.01	178
Lyndon B. Johnson	0.92	0.05	0.03	0.02	115

*Recorded by FJC as of April 3, 2012

Table 6: Racial/ethnic and gender distribution of judicial nominees by President (Johnson through Obama administrations)

the bias would come from the Senate Judiciary Committee being more likely to reject “Not Qualified” white or male nominees while pushing forward “Not Qualified” minority or female nominees, perhaps due to concerns about diversity and/or not wishing to appear biased. The observable implications of both would be that the ratings awarded to confirmed candidates by the ABA would appear skewed against women or minority candidates, even though there would be no bias associated with the ratings process itself.

George W. Bush Nominees. As noted, George W. Bush declined to allow the ABA to evaluate presumptive nominees in advance of their nominations ([Gonzales, 2001](#)). Thus, during the years 2001 to 2008, we have *all* of ABA scores awarded, which avoids the selection bias problem present elsewhere in the data. A politically awkward situation is therefore empirically quite useful.

Because only 18 African American, 27 Hispanics, and 54 women judges were nominated during the Bush II years (Table 6), I use parametric methods instead of matching. Table 7 shows results from a logit regression including race, gender, and a variety of professional and educational characteristics⁹ where the outcome variable is whether the nominee (here,

⁹I do not include the judicial common scores measuring ideology because about half are missing and

the actual nominee) received either (1) high rating (“Well Qualified”), or (2) a low rating (“Qualified” or “Not Qualified”). I include one model with dummy variables for the district of origin and one model with race and gender interacted. The results are only fleetingly significant, owing to the small number of racial/ethnic minorities and women. However, the results from the Bush II years are largely consistent with the results seen before: although not significant, the model coefficients are suggestive of African American, Hispanics, and women nominees being less likely than, respectively, whites and men to receive the higher two ABA ratings. For women and for Hispanics the effects are no longer significant; for African Americans, they are significant when the effect is allowed to vary across district court jurisdiction (Model 2).

Artificially constructed presumptive candidates. The fact that George W. Bush nominated only 18 African Americans to the district courts hampers the ability to extract meaningful estimates about his terms. To provide additional context, I therefore artificially replicate the possible pool of presumptive candidates. Using the fact that we know the rough number (if not the identities) of the presumptive candidates rejected by the ABA, I include in the data generated observations designed to present the worst possible scenario for the key results.

To create the artificial set of observations, I generated several “presumptive nominees” per President. I did so by assuming that 8% of each President’s nominees had their candidacies fail at the pre-clearance stage. (The exception here is George W. Bush.) This is significantly higher than the actual number, which appears to be around 2-4% (Lott, 2001), but closer to Barack Obama’s very high average of 7.6% (Savage, 2011). The most bias would be introduced when Presidents fail to move forward poorly rated whites: not moving these individuals forward (while moving forward poorly rated minorities and women) would result including them could potentially introduce bias.

	Model 1	Model 2	Model 3
(Intercept)	-4.83*	15.51	14.60
	(1.49)	(5318.04)	(5321.65)
Age	0.10*	0.07	0.08
	(0.03)	(0.04)	(0.05)
African American	-0.32	-2.18*	-2.36
	(0.63)	(0.99)	(1.24)
Hispanic	-0.03	-0.77	-0.08
	(0.53)	(0.79)	(0.95)
Female	-0.25	0.00	0.52
	(0.38)	(0.62)	(0.81)
Top 25 Law School	-0.00	-0.05	0.08
	(0.36)	(0.58)	(0.60)
Private Law School	0.61	0.37	0.43
	(0.32)	(0.54)	(0.55)
Law Clerk	-0.20	-0.82	-0.76
	(0.34)	(0.63)	(0.64)
Law Professor	0.55	-0.01	0.12
	(0.87)	(1.50)	(1.53)
Private Practice	0.25	-0.65	-0.48
	(0.48)	(0.84)	(0.91)
US Attorney	-0.21	-0.13	-0.25
	(0.57)	(1.17)	(1.18)
Assistant US Attorney	1.43*	1.92*	1.99*
	(0.43)	(0.75)	(0.78)
Justice Department	-0.86	-1.07	-0.88
	(0.68)	(1.44)	(1.50)
Public Defender	0.65	1.46	1.89
	(0.83)	(1.77)	(1.96)
Federal Magistrate	0.82	1.18	1.20
	(0.49)	(0.85)	(0.85)
State Judge	0.10	0.04	0.08
	(0.33)	(0.54)	(0.55)
African American*Female			0.51
			(1.98)
Hispanic*Female			-1.94
			(1.39)
District Dummies		✓	✓
N	257	257	257
$\log L$	-92.41	205.96	213.17

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 7: Logit regression results, George W. Bush nominees. Outcome variable is receiving a high ABA rating.

in a skewed post-selection sample. Thus, I initially create an artificial sample of 120 “failed nominees” who are both white, young, and poorly qualified by the ABA, and I assign them

	Model 1	Model 2	Model 3
African American	-0.62*	-0.70*	-0.74*
	(0.19)	(0.20)	(0.23)
Hispanic	-0.53*	-0.73*	-0.75*
	(0.23)	(0.24)	(0.27)
Female	-0.14	-0.31*	-0.34*
	(0.15)	(0.16)	(0.17)
African American:Female			0.15
			(0.44)
Hispanic*Female			0.11
			(0.52)
President Dummies		✓	✓
N	1749	1749	1749
$\log L$	-1005.10	-946.22	-940.15

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 8: Logit coefficients generated when including 120 generated observations to reflect unknown “presumptive nominees.” Controls for professional experience, age, and education not shown.

those covariates least linked with higher ABA ratings (including no prior judgeships, law clerkships, or private practice experience).

After including them in with the original data, I re-ran the key analyses, which are presented in Table 8. The original results are insensitive to their inclusion, particularly for African Americans and for Hispanics, for whom the relationship to high ABA scores is still negative and significant under any model specification. For women, the results are still significant once we allow the effect to vary across Presidential administration (Model 2). The results are therefore not broken, even under these extreme assumptions. In addition, incrementally increasing the number of “presumptive judges” in the artificially created set (dropping covariates), shows that the fraction of presumptive nominees dropped due to the selection bias out of the total number of nominees would have to be 10% to break the results for women, 11% to break the results for Hispanics, and 15% to break the results for African Americans – nearly twice the rate recently reported for the Obama administration (Savage,

2011). Thus, we have little evidence that the results seen here are driven by this particular selection bias problem.

I note that these sensitivity tests do not yet address the second source of bias identified – actual nominees whose nominations were withdrawn. These include around 200 publicly named nominees whose candidacies were (1) withdrawn by the White House or by the nominees themselves, (2) rejected by the Senate, or (3) withdrawn due to the nominee’s death. For these individuals, I have begun collecting the same information as those who were eventually confirmed and invested – e.g., data on their legal education, previous professional experience, any judicial positions (if any), and their ABA rating. The data were not ready by time of writing, but their relatively small number (200 compared to 1,600) means that the bias resulting from their exclusion is likely limited, and a preliminary check is provided by the same analysis presented in Table 8.

7 ABA Ratings and Party Bias

A remaining issue is whether, as many have alleged, there is an ideological bias to the scores assigned by the American Bar Association, one that discriminates against ideological conservatives and Republicans. Here, at least three empirical studies – [Smelcer, Steigerwalt and Vining Jr \(2011\)](#); [Lott \(2001\)](#); [Lindgren \(2001\)](#) – have found such an effect looking at the U.S. Courts of Appeals. Although there is good reason to think that appointments to the U.S. Courts are more likely to be driven by political concerns, a politically biased vetting process would likely also extend to the hundreds more appointments to the U.S. trial courts.

To test this, I focus on two measures of partisanship. The first is the party of the appointing president, while the second is the nominee’s judicial common score. The judicial common score takes advantage of “senatorial courtesy,” the longstanding practice of Presidents to consult U.S. Senators on judicial vacancies in their home states. Thus, the judicial

common scores are based on either the (1) the ideological common score of the appointing president or, in instances where the President and the senior senator of the judge’s home state are the same, (2) the common score of the senior senator. In addition, because the sentiment of discrimination comes largely from Republican administrations, I more closely analyze judges confirmed during those time periods.

As before, I match on numerous variables: race, gender, age (broken into cohorts), state judge, U.S. Attorney or Assistant U.S. Attorney, previous judicial experience (as a state judge or term-limited federal judge), law school rank (and whether the law school was private or public), clerkship experience, private practice experience, law professor, and public defender. Because the purpose of this analysis is to detect differences across partisan appointments, I do not match on the identity of the President making the appointments. I also do not match on the judicial common score. (Characteristics of the post-matched population is provided by Table 3.) As before the outcome variable of interest is receiving a high rating (“Exceptionally Well Qualified” or “Well Qualified”) from the ABA.

Post-matching logit results are presented in Table 9. I include at four specifications: looking at party of the appointing president (Model 1), ideology (Model 2), both (Model 3), and both interacted (Model 4, which I use to test the idea that ideologically conservative nominees might be awarded lower scores by the ABA, but mostly when they are appointed by Republican administrations). As the results in Table 9 show, however, there is no statistically significant relationship between any of the possible treatments and receiving a high or low ABA score. For party of the appointing President, the relationship is substantively weak and falls shy of statistical significance (despite the substantial post-matching sample size), and this result dovetails with the parametric regression results presented in Figure 1 as well as Table 11. The same is true for looking at ideology: we simply cannot rule out that there is no relationship between more conservative ideology and receiving a lower score. An interaction between the two is also not significant. Thus, I see no evidence that candidates nominated

	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.43*	0.35*	0.46*	0.45*
	(0.11)	(0.07)	(0.15)	(0.22)
Republican	-0.08		-0.21	-0.20
	(0.14)		(0.26)	(0.28)
Judicial Common Score		-0.06	0.19	0.13
		(0.21)	(0.37)	(0.70)
Judicial Common Score * Republican				0.08
				(0.82)
<i>N</i>	861	781	781	781
AIC	1175.27	1083.71	1084.81	1086.83
BIC	1213.34	1120.99	1140.74	1161.40
log <i>L</i>	-579.64	-533.85	-530.40	-527.41

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 9: Effect of party of appointing presidents, post matching. Logit coefficient estimates, with outcome being receiving a high ABA score. Larger (positive) judicial common scores indicate increased conservative ideology.

by Republican presidents or who are more conservative are systematically awarded lower scores than similarly pedigreed candidates nominated by Democrat presidents or who are more liberal. This is a finding that stands in contrast with earlier literature on appointments made to the Courts of Appeals, an aspect which I discuss below.

8 ABA Ratings as Predictors of Judicial Performance

The results presented so far call into question the impartiality of American Bar Association qualification ratings for African Americans and for women, while complicating earlier conclusions about their partisanship or ideological bias. I now turn to a separate question, which concerns the utility of ABA scores. Given how much effort goes into calculating these scores, and deference paid to them by political actors, we would expect that ABA ratings serve some useful function or signal. In the nominations context, the greatest utility would be if ABA scores somehow predict how judge will fare once invested onto the bench – that

is, how frequently the cases they write are reversed or upheld.

I note at this point that simple reversal is not an universally agreed-upon measure of judicial “quality” or “performance,” which are inherently slippery concepts, and that a lively normative debate is ongoing about whether, and to what extent, judges should be held to performance standards. When it comes to the lower courts, however, there is some agreement that certain judges systematically produce opinions of poorer quality or poorer legal reasoning, which in turn are more consistently reversed. (This assertion might be ring less true in the higher courts, where the discretionary nature of review means that high-quality opinions may be turned over due primarily to political considerations.) Thus, if ABA scores are useful predictors of anything, it should be of a district judge’s reversal rate.

Case outcomes data. To test this possibility, I use an extant database of cases by [Songer, Kuersten and Haire \(2007\)](#). These cases represent a randomly selected subset of 12,519 cases appealed from the U.S. District Courts to the U.S. Courts of Appeals between 1960 and 2002.¹⁰ For each case, I have data on the ultimate decision by the appeals panels. I operationalize this as being dichotomous: the appeals panel either upholds the lower-court opinion or reverses it, either in its entirety or in part. (More sophisticated measures of higher-court outcomes do not meaningfully affect the results.) In addition, I also have the identity of the U.S. District Court judge who wrote the lower-court opinion, including his or her ABA rating. I use both pieces of information information to calculate for 1,044 district court judges his or her reversal rate over this time period. (Again, I limit the population to judges who were confirmed after 1960.) A histogram of the distribution of judges’ reversal rates is presented in [Figure 2](#).

Results. [Table 10](#) presents the results from a simple OLS model where the key explanatory variable is the score awarded to the judge, with the outcome become the judges’ reversal rate

¹⁰Data more recent than 2002 were unavailable at the time of writing.

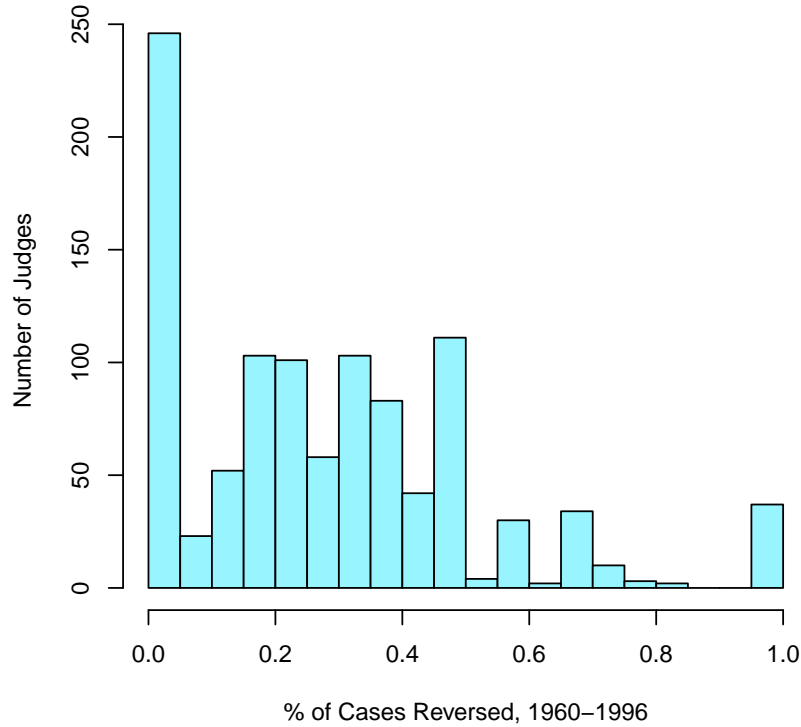


Figure 2: Reversal rates for U.S. District judges confirmed after 1960 (using cases decided on appeal from 1960 to 2002).

on cases decided between 1960 and 2002. I either dichotomize the ABA ratings categories (“highly qualified” versus “poorly qualified,” Model 1) or include the full spectrum of ABA ratings, taking “Not Qualified” as the baseline (Model 2). Thus, Model 2 explicitly tests whether “not qualified” candidates are indeed substantively different than those the ABA deems “above the bar.” What Table 10 demonstrates, however, is that ABA ratings under neither schema are predictive of a judge’s reversal rate. Indeed, not only are the coefficients close to zero and statistically insignificant, but the R^2 of the regression is close to zero as well.¹¹

¹¹Stylizing the outcome variable not as the judge’s overall reversal rate, but as the probability that *an individual case* will be reversed (in whole or in part) also results in null results with regard to the lower court

	Model 1	Model 2	Model 3
(Intercept)	0.30*	0.26*	0.40*
	(0.01)	(0.09)	(0.05)
High ABA Rating (yes or no)	-0.02		-0.01
	(0.01)		(0.02)
“Qualified”		0.04	
		(0.09)	
“Well Qualified”		0.02	
		(0.09)	
“Exceptionally Well Qualified”		0.06	
		(0.10)	
Republican			-0.03*
			(0.02)
Law Clerk			0.01
			(0.02)
State Judge			0.02
			(0.02)
Top 25 Law School			-0.03*
			(0.02)
Circuit Dummies			✓
<i>N</i>	1044	1044	1043
<i>R</i> ²	0.00	0.00	0.08
adj. <i>R</i> ²	0.00	-0.00	0.04
Resid. sd	0.24	0.24	0.24

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 10: OLS regression of a judge’s reversal rate (for cases decided on appeal between 1960-2002) on ABA qualification ratings.

Because ABA scores ostensibly collapse the entirety into nominee’s qualifications into one easy-to-digest value, they should be predictive of judges’ reversal rates without additional controls. Nonetheless, it is likely that norms about reversal vary across jurisdiction and through time, which could reveal the ABA scores’s predictive value. It is also possible that reversal varies independent of ABA score according to law school rank, previous judicial experience, etc., and that the scores provide additional value added on top of these other judge’s ABA score when judge-specific random effects are included along with dummy variables for Circuit or year.

kinds of signals. I therefore include additional predictors in the analysis (Model 3), including dummy variables representing the judicial district in which the appeal arose (“Circuits”). The results from this specification (Table 10, Model 3), show that there is still no relationship between ABA scores awarded to a judge and his or her rate of reversal. On the other hand, we do see a negative relationship between graduating from a Top 25 top law school and reversal rate, as well as a negative relationship between being Republican and being reversed. Dummy variables for the U.S. Appeals Courts are also significant (not shown), which suggests that norms about reversal may vary from jurisdiction to jurisdiction. In sum, there is little evidence that the score received by a judicial nominee in any way predicts how successful he or she will be in avoiding reversal.

9 Conclusion

The contributions of this paper are threefold. First, the results show no differences between nominations made by Democrats and Republicans or among nominees with different ideological common scores – a finding in contrast to previous literature. Although more work needs to be done, the reason may be rooted in the different roles played by U.S. District and U.S. Appeals courts. The latter perhaps have a stronger reputation for being partisan; not only are their nominations taken more seriously by political actors, but decision making on the appeals courts has been shown to be closely linked with party affiliations (Sunstein et al., 2006). Thus, it would not be surprising to see partisanship play a key role in the nominations of appeals court judges and less so for judges at the district level.

Second, although the results show no differences between Democrat and Republican nominees, my findings suggest that women and minority judicial candidates systematically receive lower qualification ratings from the ABA. This is the case both a priori and also when using matching or other controls to compare candidates who are similar or identical

across key professional, educational, and political characteristics. The results also appear robust to potential selection issues arising from the practice of Presidents “pre-clearing” potential nominees with the ABA. The effect is present both in Republican and Democrat administrations, and sensitivity analyses suggest that it is not being driven by variables omitted from the analysis.

One way to understand these results is that the law is a prestige-oriented profession – one driven by high-status accomplishments and the general appearance of success. To this extent, it is not surprising that law school rank, previous legal clerkship experience, private practice experience, and public defender experience are predictive of the kind of ABA rating a nominee will receive. On the other hand, in instances where prestige, power, and appearances of success are paramount, we might also not be surprised that women and minorities may be systematically disadvantaged.

Third, and perhaps most importantly, the findings show that ABA ratings are not predictive of judges’ ultimate performance once they are confirmed. Indeed, the analysis here demonstrates that nominees designated as “Not Qualified” to serve by the ABA have reversal rates that differ little from those awarded the stellar “Exceptionally Well Qualified” and “Well Qualified” ratings. This fact is surprising given that the ABA ostensibly takes into account those aspects which would make for a strong judicial career – both objective criteria like law school attended, and also subjective criteria such as “temperament,” “competence,” and “integrity.”

Ultimately, however, the results presented here call into question the substantial reliance by Presidential administrations, by the Department of Justice, by Senate committees, and by the media on ratings produced by the ABA and by similar organizations. Since the Eisenhower Administration, the ABA has enjoyed the privilege of “pre-clearing” the list of presumptive nominees put forth by the White House, and journalistic evidence suggests that dozens of candidates preliminary deemed “not qualified” by the ABA have had their candi-

dacies derailed. Here, I have presented evidence that this rating process could be resulting in systematic bias against women and minorities. I have also presented evidence that these ratings are not particularly useful in terms of predicting long-term judicial performance. Taken together, they suggest that continuing to allow non-governmental organizations like the ABA to “pre-clear” nominees provides little benefit.

Appendix

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	-50.01*	-49.67*	47.95	47.67	58.63
	(9.16)	(11.00)	(68.42)	(68.51)	(74.80)
Age	0.06*	0.06*	0.06*	0.06*	0.06*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
African American	-0.81*	-0.99*	-0.98*	-0.94*	-1.18*
	(0.20)	(0.22)	(0.22)	(0.24)	(0.26)
Hispanic	-0.74*	-0.64*	-0.65*	-0.72*	-0.77*
	(0.23)	(0.27)	(0.27)	(0.30)	(0.36)
Female	-0.41*	-0.50*	-0.50*	-0.51*	-0.60*
	(0.16)	(0.18)	(0.18)	(0.19)	(0.21)
Republican	-0.04	-0.14	-0.04	-0.04	-0.48
	(0.11)	(0.23)	(0.36)	(0.36)	(0.43)
Top 25 Law School	0.31*	0.33*	0.34*	0.34*	0.13
	(0.11)	(0.12)	(0.12)	(0.12)	(0.15)
Private Law School	0.11	0.17	0.17	0.17	0.03
	(0.11)	(0.12)	(0.12)	(0.12)	(0.14)
Law Clerk	0.20	0.20	0.20	0.20	0.22
	(0.14)	(0.16)	(0.16)	(0.16)	(0.17)
Law Professor	-0.04	-0.02	-0.01	-0.01	-0.01
	(0.23)	(0.24)	(0.24)	(0.24)	(0.26)
Private Practice	0.51*	0.49*	0.52*	0.52*	0.50*
	(0.20)	(0.22)	(0.23)	(0.23)	(0.24)
US Attorney	-0.23	-0.25	-0.23	-0.23	-0.23
	(0.20)	(0.21)	(0.21)	(0.21)	(0.23)
Assistant.US.Attorney	0.63*	0.63*	0.63*	0.62*	0.66*
	(0.15)	(0.16)	(0.16)	(0.16)	(0.18)
Justice Department	0.13	0.16	0.15	0.15	-0.06
	(0.26)	(0.30)	(0.30)	(0.30)	(0.33)
Public Defender	0.20	0.50	0.54	0.55	0.38
	(0.28)	(0.32)	(0.32)	(0.32)	(0.35)
Federal Magistrate	0.79*	0.66*	0.64*	0.64*	0.72*
	(0.22)	(0.24)	(0.24)	(0.24)	(0.26)
Federal Bankruptcy	0.82	0.88	0.89	0.91	0.91
	(0.48)	(0.49)	(0.49)	(0.49)	(0.55)
State Judge	0.27*	0.27*	0.27*	0.27*	0.17
	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)
Commission Year	0.02*	0.02*	-0.03	-0.03	-0.03
	(0.00)	(0.01)	(0.03)	(0.03)	(0.04)
ideology		0.04	-0.09	-0.08	0.47
		(0.32)	(0.33)	(0.33)	(0.45)
African American*Female				-0.21	-0.10
				(0.56)	(0.58)
Hispanic*Female				0.32	0.36
				(0.63)	(0.67)
Presidential Administration Dummies			✓	✓	✓
District Court Dummies					✓
<i>N</i>	1629	1421	1421	1421	1421
$\log L$	-967.73	-843.26	-818.16	-811.95	-476.73

Standard errors in parentheses
 * indicates significance at $p < 0.05$

Table 11: Logit regression results, judges nominated between 1996 and 2002. Outcome variable is receiving a high ABA rating.

References

American Bar Association. 2009. “Standing Committee on the Federal Judiciary: What It is and How it Works.”.

URL: http://www.americanbar.org/content/dam/aba/migrated/2011_build/federal_judiciary/federal_jud
4, 5, 9, 20

Boyd, C.L. 2011. “Federal District Court Judge Ideology Data.”.

URL: <http://cLboyd.net/ideology.html> 13

Epstein, L., A.D. Martin, J.A. Segal and C. Westerland. 2007. “The judicial common space.” *Journal of Law, Economics, and Organization* 23(2):303. 13

Giles, M.W., V.A. Hettinger and T. Peppers. 2001. “Picking federal judges: A note on policy and partisan selection agendas.” *Political Research Quarterly* 54(3):623–641. 13

Gonzales, A. 2001. “Letter from Alberto Gonzales, U.S. Attorney General, to Martha Barnett, President of the American Bar Association.”. 6, 21

Greiner, D.J. and D.B. Rubin. 2010. “Causal Effects of Perceived Immutable Characteristics.” *Review of Economics and Statistics* . 13

Ho, D.E., K. Imai, G. King and E.A. Stuart. 2007. “Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference.” *Political Analysis* . 11

Iacus, S., G. King and G. Porro. 2009. “CEM: Software for Coarsened Exact Matching.”. 12

Iacus, S.M., G. King and G. Porro. 2011. “Multivariate Matching Methods That are Monotonic Imbalance Bounding.” *Journal of the American Statistical Association* . 12

Just the Beginning Foundation. 2012. “Integration of the Federal Judiciary.”.

URL: <http://www.jtbf.org> 2

Keele, L. 2010. “An Overview of `rbounds`: An R package for Rosenbaum Bounds Sensitivity Analysis with Matched Data.” *White Paper. Columbus, OH* pp. 1–15. 18

Lindgren, J.T. 2001. “Examining the American Bar Association’s Ratings of Nominees to the US Courts of Appeals for Political Bias, 1989-2000.” *Northwestern Law Legal Working Paper Series* p. 37. 2, 5, 8, 25

Lott, J. 2001. “American Bar Association, Judicial Ratings, and Political Bias, The.” *Journal of Law & Politics* 17:41. 2, 5, 6, 8, 20, 22, 25

Lott, J. 2006. “Pulling Rank.” *The New York Times* .

URL: <http://www.nytimes.com/2006/01/25/opinion/25Lott.html> 6

Mirengoff, P. 2010. “The American Bar Association Exposes Its Liberal Bias Once Again.”.

URL: <http://www.powerlineblog.com/archives/2010/02/025696.php> 6

- Poole, K.T. 1998. "Recovering a basic space from a set of issue scales." *American Journal of Political Science* pp. 954–993. 13
- Raben, R. 2011. "The ABA Ratings and Minority Nominees: Shedding Light on Disparate Impact." .
URL: www.acslaw.org/acsblog/the-aba-ratings-and-minority-nominees-shedding-light-on-disparate-impact 7
- Rosenbaum, P.R. 2002. *Observational studies*. Springer Verlag. 18
- Saks, M.J. and N. Vidmar. 2001. "Flawed Search for Bias in the American Bar Association's Ratings of Prospective Judicial Nominees: A Critique of the Lindgren Study." *Journal of Law & Politics* 17:219. 5
- Savage, C. 2011. "Ratings Shrink President's List for Judgeships." . 2, 3, 6, 7, 20, 22, 24
- Sen, M. and O.T. Wasow. 2011. "Reconciling Race and Causation: Methods to Extract Meaningful Causal Inferences About Race." .
URL: http://scholar.harvard.edu/msen/files/sen_wasow_causality.pdf 13
- Smelcer, S.N., A. Steigerwalt and R.L. Vining Jr. 2011. "Bias and the Bar: Evaluating the ABA Ratings of Federal Judicial Nominees." *Political Research Quarterly* . 2, 5, 6, 8, 25
- Songer, Donald, Ahslyn K. Kuersten and Susan B. Haire. 2007. "The United States Courts of Appeals Database." . 28
- Sunstein, C.R., D. Schkade, L.M. Ellman and A. Sawicki. 2006. *Are Judges Political?* Brookings Institution Press. 31
- Tetreault, S. 2010. "Reid Criticizes Lawyers Group." *Law Vegas Review-Journal* . 7
- Whelan, E. 2010. "Re: The ABA and Ninth Circuit Nominee Goodwin Liu." *The National Review Online* .
URL: <http://www.nationalreview.com/bench-memos/49265/re-aba-and-ninth-circuit-nominee-goodwin-liu/ed-whelan> 6
- Zuk, Gary, Deborah J. Barrow and Gerard Gryski. 2009. "Multi-User Database on the Attributes of United States District Court Judges, 1801-2000." ICPSR 4553, Ann Arbor, MI: Inter-university Consortium for Political and Social Research. 20