Missouri Law Review

Volume 76
Issue 2 *Spring 2011*

Article 2

Spring 2011

Distinguishing Judges: An Empirical Ranking of Judicial Quality in the United States Court of Appeals

Robert Anderson

MISSOURI LAW REVIEW

VOLUME 76 SPRING 2011 NUMBER 2

ARTICLES

Distinguishing Judges: An Empirical Ranking of Judicial Quality in the United States Courts of Appeals

*Robert Anderson IV**

ABSTRACT

This Article presents an empirical performance ranking of 383 federal appellate judges who served on the United States Courts of Appeals between 1960 and 2008. Like existing judge evaluation studies, this Article uses citations from judicial opinions to assess judicial quality. Unlike existing citation studies, which treat positive and negative citations alike, this Article ranks judges according to the mix of positive and negative citations to the opinions, rather than the number of citations to those opinions. By distinguishing between positive and negative citations, this approach avoids ranking judges higher for citations even when the judges are being cited negatively. The results are strikingly different from those found in the existing citation count-based studies of judicial performance. When the mix of positive and negative citations is taken into account, many of the most highly cited judges from the citation-count studies are only average and some of the average judges in the citation-count studies emerge as the most positively cited. The results suggest there is an objective performance measure that can measure judicial performance and provide incentives for fidelity to the rule of law.

* Associate Professor of Law, Pepperdine University School of Law. J.D., New York University; Ph.D., Stanford University. I would like to thank Stephen Choi, Frank Cross, Mitu Gulati, and Robert Pushaw for helpful comments on a previous version of this Article and Ryan Griffee for excellent research assistance. I should disclose that one of the judges evaluated in this Article, Kenneth W. Starr, was the Dean of Pepperdine University School of Law while this Article was written, and another, Deanell Reece Tacha, will begin service as Dean on June 1, 2011.

I. INTRODUCTION

How can one evaluate the performance of federal appellate judges? This question implicitly arises every time a federal appellate judge is nominated to the United States Supreme Court. And because the federal appellate bench is the most common source of Supreme Court nominees in recent decades,[1] this question is relevant to most modern Supreme Court nominations. But the question of judicial performance is at least as important outside the context of Supreme Court appointments, as the courts of appeals are the final arbiters of most disputes in the federal courts. Thus, the outcome of virtually every litigated matter in the federal system hinges on the quality of federal appellate decision-making, and therefore the performance of these judges implicates fundamental questions about the rule of law.

However, the importance of evaluating the performance of federal judges has not motivated systematic assessment of individual judges' work product in legal scholarship. Indeed, aside from anecdotal information, little is known about the performance of individual federal appellate judges. Of course there is no dearth of scholarly critique of federal courts' *products* – the opinions in individual cases – but these critiques are not systematically organized into evaluation of the *producers* of the opinions – the judges themselves. Thus, in spite of the fact that the performance of individual judges has important implications for the functioning of the judicial system and rule of law, scholars still do not have a good idea of which judges are performing well and which judges are performing poorly. Few academic studies have even attempted to evaluate federal judges using quantitative data, and, when they have, they have generally received harsh criticism from scholarly commentators.[2]

The recent nomination of then-Judge Sonia Sotomayor to the Supreme Court illustrates the de facto alternative to systematic approaches to judicial quality. The Sotomayor nomination, like nominations of federal appellate judges in the past, tended to focus on detailed scrutiny of a small number of high-profile opinions, distracting from the broader, systematic examination of the nominee's body of work as a whole.[3] In the absence of reliable information about judicial performance, center stage in the debate is yielded to anecdotal accounts of anonymous sources,[4] isolated remarks from the judge's

---

1. *See* Lee Epstein et al., *The Norm of Prior Judicial Experience and Its Consequences for Career Diversity on the U.S. Supreme Court*, 91 CAL. L. REV. 903, 908 (2003) (noting that "presidents now look primarily to the U.S. courts of appeals to identify potential nominees" to the Court).

2. *See infra* Part V.A.

3. *See GOP Homes in on Controversial Sotomayor Speeches*, CNNPOLITICS.COM, July 19, 2009, http://edition.cnn.com/2009/POLITICS/07/16/sotomayor.hearing/index.html.

4. *See, e.g.*, Jeffrey Rosen, *The Case Against Sotomayor*, NEW REPUBLIC, May 4, 2009, *available at* http://www.tnr.com/politics/story.html?id=45d56e6f-f497-4b19-

public appearances, and short passages in opinions culled from the tens of thousands of pages the nominee has written. Although we have the benefit of a more thorough evaluation from the American Bar Association, its approach has been called biased[5] and may be no more objective than the confirmation hearings. The result is that the evaluation of judicial performance is biased, subjective, and based on a narrow slice of information rather than on the judge's record as a whole.

The frustration with the prevailing approaches to assessing judicial quality, both in the context of Supreme Court appointments and otherwise, has led scholars and legal commentators to develop quantitative techniques to measure judicial performance.[6] The most prominent approaches in recent years use large databases of citations to evaluate the "influence," "prestige," or "quality" of judges. One of the first such papers, by Professors Landes, Lessig, and Solimine (hereinafter "Landes et al."), used citation counts to opinions to measure "judicial influence" in the federal courts of appeals.[7] More recently, Professors Choi and Gulati have expanded on the Landes et al. study, using citation counts to measure "productivity," "quality," and "independence" on the federal courts of appeals.[8] In contrast to the typical evaluation of judges' opinions that legal scholars perform in law reviews, the citation literature abstracts away from the details of the cases to systematically evaluate the whole body of the judges' work product.

The citation studies have revealed information about judicial performance that was not previously well known outside the ranks of experienced appellate advocates and federal judges themselves – if it was known at all. Perhaps for this reason, the studies have attracted considerable attention in legal scholarship, including multiple responses to Choi and Gulati's first

---

9c63-04e10199a085 (quoting an anonymous former Second Circuit clerk describing Judge Sotomayor as "not that smart and kind of a bully on the bench").

5. *See, e.g.*, James Lindgren, *Examining the American Bar Association's Ratings of Nominees to the U.S. Courts of Appeals for Political Bias, 1989-2000*, 17 J.L. & POL. 1, 26 (2001). For a critical response, see Michael J. Saks & Neil Vidmar, *A Flawed Search for Bias in the American Bar Association's Ratings of Prospective Judicial Nominees: A Critique of the Lindgren Study*, 17 J.L. & POL. 219 (2001). *See also* Scott Baker et al., *The Continuing Search for a Meaningful Model of Judicial Rankings and Why It (Unfortunately) Matters*, 58 DUKE L.J. 1645, 1655 n.35 (2009); Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. CAL. L. REV. 23, 35-36 (2004); Steven G. Gey & Jim Rossi, *Empirical Measures of Judicial Performance*, 32 FLA. ST. U. L. REV. 1001, 1001-02 (2005).

6. *See, e.g.*, Choi & Gulati, *supra* note 5, at 26 (explaining that their "frustration with the current appointment process" inspired their judge ranking project).

7. William M. Landes et al., *Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges*, 27 J. LEGAL STUD. 271 (1998).

8. Choi & Gulati, *supra* note 5, at 33.

judge ranking paper in the *Southern California Law Review*,[9] a symposium published in the *Florida State University Law Review*,[10] and conference proceedings published in the *Duke Law Journal*.[11] As might be expected, however, the responding scholars and judges have not enthusiastically welcomed this quantitative intrusion into the traditional purview of qualitative legal commentary. Although some of the commentators' responses have focused on very fair criticisms of the methodology employed by Choi and Gulati, several scholarly responses seem to have a broader point – one that rejects the very idea of quantitative assessment of judging.[12]

Those who reject the notion that judicial performance is quantifiable will, of course, be disappointed by any variation of the Choi and Gulati techniques. But for those who see quantitative methods as a valuable tool in systematic assessment of the judiciary, there is an opportunity to build on existing citation studies to reveal another perspective on judicial performance. The opportunity arises because of two key limitations of the citation studies that limit the effectiveness of the existing work. First, the citation studies are "count-based," meaning that the number of citations is the key variable of interest in evaluating the judges. Among other problems, this approach treats negative citations the same as positive citations, even though negative citations might reflect negatively on judicial quality. Second, the citation studies rate judges based only on opinions they have authored, rather than all cases in which they have participated. Opinion authorship, although closely tied to individual judges, raises a host of problems, not the least of which is selection bias in opinion assignment. Constructing a judicial evaluation technique that responds to these two problems offers a clearer, more comprehensive view of judicial performance.

This Article attempts to construct such a technique for ranking federal appellate judges – one that does not have the same drawbacks of the existing citation-count studies. The problem of treating positive and negative citations alike is addressed by using "treatment" data from Shepard's Citations (provided by LexisNexis) to rank judges according to the positive or negative citations of their peers: other federal appellate judges. Using treatment data allows judges whose decisions are cited more positively to receive higher rankings and judges who are cited more negatively to receive lower rankings. The problem of selection bias is addressed by using panel membership, rather than opinion authorship, as the link between judges and citations. This means

9. *See, e.g.*, William P. Marshall, *Be Careful What You Wish For: The Problems with Using Empirical Rankings to Select Supreme Court Justices*, 78 S. CAL. L. REV. 119 (2004); Workshop on Empirical Research in the Law, *On Tournaments for Appointing Great Justices to the U.S. Supreme Court*, 78 S. CAL. L. REV. 157 (2004).

10. Symposium, *Empirical Measures of Judicial Performance*, 32 FLA. ST. U. L. REV. 1001 (2005).

11. Jeffrey M. Chermerinsky & Jonathan L. Williams, *Measuring Judges and Justice*, 58 DUKE L.J. 1173 (2009).

12. *See infra* notes 43-44 and accompanying text.

that judges who contribute to producing a decision that is positively cited when they did not author the opinion are ranked higher than those who only contribute to producing a decision that is positively cited when they author the opinion. These two innovations are designed to produce a measure of the *quality of the average opinion* produced jointly by an appellate panel, rather than the visibility or notoriety of opinions authored individually by a judge, as in the citation-count measures.

Although this project was conceived as a means of building on the existing citation-count studies, the results are so strikingly different from those of the citation-count models that this study is more properly viewed as a break with the existing literature. Indeed, the picture of judicial performance that emerges from this study poses a challenge to widely-held conceptions about the identities of the "top" judges in the federal appellate courts. Some of the most prominent judges in the citation-count models and law review literature appear only average when ranked by the mix of positive and negative citations to their opinions. Similarly, some relatively low-profile judges who rarely make the pages of law review scholarship emerge as some of the nation's most highly rated judges in this ranking. The reason is relatively clear: while citation counts tend to reward the most provocative judicial entrepreneurs, this study rewards the careful judicial craftsperson. Thus, the results of this study provide a means of assessing the quality of the typical decision rendered by an appellate judge, rather than the notoriety of his or her high-profile decisions.

This Article outlines a judicial evaluation tool that is as transparent and objective as in the method used by Choi and Gulati, but one that more directly measures the characteristics most people – especially litigants in the federal courts – are likely to think of and care about as judicial "quality." The justification for this alternative measure is that the average litigant likely cares much more about the quality of federal appellate judges than about the judges' passing of an ideological litmus test, or the likelihood the litigant's dispute will be immortalized in casebooks and law review articles. Moreover, this approach provides a means of assessing judicial performance for the purpose of judicial administration that is complementary to, rather than duplicative of, the productivity measures already used in judicial assessment, such as caseloads and backlogs. Finally, by using a much larger dataset and more detailed information than the existing studies, the quality measures in this study offer a preliminary but revealing look into the interplay of ideology and precedent in the federal appellate courts.

This Article proceeds as follows: Part II surveys the burgeoning literature evaluating judicial performance and explains the contribution of this Article in extending that literature. It outlines the theory that underlies the performance measurement strategy in this Article and how incorporating positive and negative citations improves that measure. Part III describes the dataset and methods used in this study. Part IV presents the results – a ranking of 383 federal appellate judges based on positive and negative citations to their decisions since 1960. This Part uses the opportunity presented by then-

Judge Sotomayor's nomination to the Supreme Court to compare her performance relative to that of the other federal appellate judges considered for the nomination. Part V applies this research to broad normative policy questions, such as the relationship between ideology and precedent and whether the Ninth Circuit should be split. Part VI concludes with the observation that both ideology and judicial quality appear to drive judicial citation patterns.

II. LITERATURE AND THEORY

A. Introduction to Existing Literature

The quantitative literature evaluating judicial performance is still in its infancy. Although evaluations based on survey responses have been around since the *Almanac of the Federal Judiciary* was first published in 1984,[13] only recently have scholars begun to use large databases to evaluate the performance of judges.[14] The quantitative work on judicial quality can be roughly grouped into two broad categories according to the study's evaluative measure.[15] One approach uses judicial outcome measures to evaluate judges, such as voting patterns and reversal rates on appeal.[16] Another group of approaches uses the number of citations to opinions from other judicial opinions, law review articles, and so forth to evaluate judges.[17] In each case, the goal is to find an objective measurement that captures something important about judicial influence, prestige, or quality, while possibly controlling for one or more variables that would potentially confound the analysis. To illustrate these two approaches, the discussion below contrasts two recent studies: the Cross and Lindquist reversal-rate study and the Choi and Gulati citation analysis study.

The reversal-rate approach is probably the most intuitive way to think about assessing performance for judges who do not sit in a court of last resort.

---

13. Perhaps the most well-known resource in this category is the ALMANAC OF THE FEDERAL JUDICIARY (1984).

14. *See, e.g.*, Landes et al., *supra* note 7.

15. There are a number of other measures that have been used for many years in judicial administration, such as caseloads, backlog statistics, etc. *See* RICHARD A. POSNER, HOW JUDGES THINK 146-47 (2008). Although these measures might be used to evaluate judges, they are typically more applicable to the district courts than to the appellate courts. *Id.* at 147.

16. The literature on voting patterns primarily has been used to measure judicial ideology, and it has been examined exhaustively in the Supreme Court context. *See* Andrew D. Martin & Kevin M. Quinn, The Dimensions of Supreme Court Decision Making: Again Revisiting the Judicial Mind (May 2, 2001) (unpublished manuscript), *available at* http://www.stat.washington.edu/quinn/papers/supctmeas.pdf. Analyzing voting patterns is less useful in the courts of appeals because there are so few dissents and therefore little disagreement to examine.

17. *See* Landes et al., *supra* note 7.

This is true in part because appellate courts themselves invite this interpretation by describing trial courts as having "erred" when they reverse the trial courts. This reversal-rate approach involves comparing the rates at which individual judges' decisions are reversed on appeal (or reversed on certiorari in the case of federal appellate judges).[18] Assuming that there is a "correct" disposition of most cases (an aggressive assumption), or at least that there are "incorrect" ways of resolving some cases (a less aggressive assumption), and that the superior court is "correct" more often than the lower court (certainly debatable, but not implausible), then the rate of reversal approach might capture a measure of judicial quality.[19] The interpretation is that lower rates of reversal, possibly as moderated by control variables, translate into higher judicial quality.

The strengths and limitations of the reversal rate approach are well illustrated by a recent study by Frank Cross and Stefanie Lindquist.[20] Cross and Lindquist examined Supreme Court review of federal appellate judges' decisions from 1989 to 2000, computing reversal rates and comparing rankings based on those rates with the findings from Choi and Gulati's citation analysis study, discussed below.[21] The authors found that highly cited judges tended to fare slightly worse at the Supreme Court level than less frequently cited judges, although the difference was not statistically significant.[22] Digging deeper into the data, however, Cross and Lindquist found that highly cited judges received more affirmances *and* reversals by the Supreme Court – in short, they tended to be reviewed by the Court more often.[23] Cross and Lindquist were cautious about their results, however, arguing that their approach "may capture only one dimension of judicial quality."[24] The authors therefore augmented their study with a cluster analysis of judicial "types," which placed judges into categories but did not purport to produce an ordinal ranking of the judges on a single scale.[25]

The reversal-rate approach has a certain appeal as a measure of performance or at least as a measure of lower court fidelity to superior court preferences. But the approach has a number of weaknesses that limit its effectiveness. One defect is that cases granted certiorari are likely a highly biased

18. *See* Frank B. Cross & Stefanie Lindquist, *Judging the Judges*, 58 DUKE L.J. 1383, 1402 (2009).

19. Of course, it is not clear that reversal rates really represent "errors" in any objective sense, except perhaps in the case of unanimous reversals, discussed in the subsequent paragraph.

20. Cross & Lindquist, *supra* note 18.

21. *Id.* at 1406-14.

22. *Id.*

23. *Id.*

24. *Id.* at 1414.

25. *Id.* at 1414-29.

sample of the work product of judges, both in terms of ideology and quality,[26] and, in any event, even if a certiorari review is granted for a case, the fact that nine justices disagree with a particular decision is not necessarily an indication of the decision's low quality.[27] The problem of ideological bias could be ameliorated somewhat by using only unanimous reversals or summary reversals, but that remedy means reducing the sample size further, leading to the next, and more significant limitation of the method. The sample size of Supreme Court review is simply too small for meaningful comparisons of individual appellate judges and possibly even too small for meaningful comparisons of whole circuits. Even the most frequently reviewed judges were not reversed or affirmed more than nine times in the Cross and Lindquist study, except for Judge Stephen Reinhardt of the Ninth Circuit, who had fourteen reversals.[28] Thus, this reversal-rate approach may be appropriate for evaluating district court judges where the problems of bias are reduced and the sample size is considerably greater, but the approach may not work well for evaluating intermediate appellate judges.[29]

The second approach to evaluating judges – the citation analysis approach – uses citations by peers, rather than reversals by superiors, as the relevant data for evaluating judicial quality.[30] The idea of using citations to evaluate judges as historical figures is not new,[31] but systematic quantitative studies of the quality of judges have emerged only in the last decade.[32] The first significant step on this path was in 1998, when Landes et al. published

---

26. Whether the cases are biased in terms of ideology or quality depends on one's theory of how the Supreme Court decides to grant certiorari. That is, does the Court grant certiorari when lower court decisions are ideologically distant from the Court or according to more traditional legal criteria, such as when the decisions are "wrongly decided" in some non-ideological sense, or when circuit splits have developed? This is exactly the type of question the results of this study can help resolve. Either way, it seems uncontroversial that the Supreme Court does not choose lower court cases randomly, so the sample of cases granted certiorari will be biased in some important way.

27. As Professor Chemerinsky put it, "[i]t is wrong to equate a reversal with a mistake by the lower court. The Supreme Court gets the final word, but that does not mean that its rulings are 'right' and reversed lower courts are 'wrong.'" Erwin Chemerinsky, *The Myth of the Liberal Ninth Circuit*, 37 LOY. L.A. L. REV. 1, 2 (2003).

28. Cross & Lindquist, *supra* note 18, at 1407-08.

29. *See* POSNER, *supra* note 15, at 147.

30. Landes et al., *supra* note 7, at 276-77.

31. *See, e.g.*, Judge Posner's writings on judicial greatness, such as RICHARD A. POSNER, CARDOZO: A STUDY IN REPUTATION (1990); Richard A. Posner, *The Learned Hand Biography and the Question of Judicial Greatness*, 104 YALE L.J. 511 (1994) (reviewing GERALD GUNTHER, LEARNED HAND: THE MAN AND THE JUDGE (1994)); *see also* Henry T. Greely, *Quantitative Analysis of a Judicial Career: A Case Study of Judge John Minor Wisdom*, 53 WASH. & LEE L. REV. 99, 101 (1996).

32. For a dated but comprehensive survey of the literature on citation analysis in studying judges more generally, see Greely, *supra* note 31, at 101-09.

their groundbreaking study of judicial influence.[33] In that study, the authors explored the advantages and disadvantages of citation analysis and ranked federal appellate judges by citation counts.[34] In the years that followed the Landes et al. study, a flurry of follow-up studies appeared, with scholars using citation analysis to study Supreme Court justices,[35] courts of appeals judges,[36] state supreme court judges,[37] and Australian judges.[38]

In this citation-analysis line of research, the most provocative recent work has come in a series of articles by Stephen Choi and Mitu Gulati.[39] Choi and Gulati's approach is similar to that of Landes et al. in that both use citation counts to evaluate judges. However, Choi and Gulati updated the analysis by using more recent data and introducing an innovation to the rankings in the form of an "independence" score.[40] But perhaps the most significant difference between the Landes et al. study and the work by Choi and Gulati is that Choi and Gulati explicitly make the normative argument that the rankings should be used to evaluate judges for promotion to the Supreme Court.[41] In brief, the authors propose that the Supreme Court nomination and confirmation process could be improved by a "tournament" in which federal appellate judges compete according to their quantitative criteria for elevation to the Court.[42] Thus, the authors go beyond the largely descriptive or theoretical work of the Landes et al. study to make aggressive normative arguments about how the rankings should be used.

The Choi and Gulati studies have generated significant response from both scholars and federal judges. Most of the response, however, has been negative, and some of it emphatically so.[43] Commentators have expressed a variety of criticisms, some specific to the Choi and Gulati project or its me-

---

33. Landes et al., *supra* note 7.

34. *Id.*

35. *See, e.g.*, Montgomery N. Kosma, *Measuring the Influence of Supreme Court Justices*, 27 J. LEGAL STUD. 333 (1998).

36. *See, e.g.*, David Klein & Darby Morrisroe, *The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals*, 28 J. LEGAL STUD. 371, 379 (1999).

37. *See, e.g.*, Stephen J. Choi et al., *Judicial Evaluations and Information Forcing: Ranking State High Courts and Their Judges*, 58 DUKE L.J. 1313, 1328-31 (2009) and works cited therein.

38. *See, e.g.*, Russell Smyth and Mita Bhattacharya, *What Determines Judicial Prestige? An Empirical Analysis for Judges on the Federal Court of Australia*, 5 AM. L. & ECON. REV. 233 (2003).

39. *See* Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CAL. L. REV. 299 (2004) [hereinafter Choi & Gulati, *A Tournament of Judges?*]; Choi & Gulati, *supra* note 5.

40. Choi & Gulati, *A Tournament of Judges?*, *supra* note 39, at 310.

41. *See id.* at 299.

42. *See id.*

43. *See, e.g.*, Lawrence B. Solum, *A Tournament of Virtue*, 32 FLA. ST. U. L. REV. 1365, 1399 (2005) (calling Choi and Gulati's proposal "a spectacularly bad idea").

thodology and some that seem directed toward the very idea of quantitative studies of the judiciary. One line of criticism expresses doubt about the project of empirical assessment of judges, arguing that it is impossible to measure judicial performance with quantitative data.[44] Another line of criticism focuses on Choi and Gulati's normative claims and argues that although judicial performance might be measurable, the proposed "tournament" would create perverse incentives or otherwise would not translate into better Supreme Court justices.[45] Because some of these criticisms raise important issues about quantitative studies of judicial performance, the discussion of these issues is deferred until Part V of this Article.

A third line of criticism relates to methodological details of the techniques used by Choi and Gulati. Some of these arguments repeat well-known criticisms of citation analysis that were extensively catalogued in the original Landes et al. study.[46] Others relate more specifically to Choi and Gulati's methodology.[47] The next section details some of these criticisms and describes how the analysis used in this study responds to the limitations of the Choi and Gulati approach and citation analysis more generally. Indeed, the approach taken in this study was motivated by many of the criticisms of the citation analysis literature.[48]

B. Drawbacks of Existing Literature

1. Positive Versus Negative Citations

The most significant drawback of the existing citation studies as measures of judicial quality is their failure to distinguish among positive, negative, and neutral citations. As explained above, the citation analysis literature evaluates judges based on the number of citations to their opinions – the "citation count" – rather than the nature of the citations – the "citation treatment."[49] The use of citation counts to evaluate judges draws upon the widespread practice of counting citations to evaluate scholarly influence and quality.[50] But there are many reasons that judges and scholars cite one another, and not all

---

44. *See* Marshall, *supra* note 9, at 129-35. The author argues, "No set of quantifiable measures of judicial excellence can be free from ideological influence or political manipulation. No set of quantifiable measures can predict who will be an outstanding Supreme Court Justice." *Id.* at 135.

45. *See infra* notes 149-59.

46. Landes et al., *supra* note 7, at 272-76.

47. *See infra* Part V.A.

48. Specifically, the literature has not "distinguished between favorable, critical, or distinguishing citations." *Id.* at 273. This study does.

49. *See supra* Part II.A.

50. For an introduction to this literature and application in the law review context, see Fred R. Shapiro, *The Most-Cited Law Review Articles*, 73 CAL. L. REV. 1540 (1985).

of them are indicative of quality of the cited work.[51] This is because although many citations are positive, some are negative,[52] and, at least in the judicial context, most are neutral, incidental, or otherwise not meaningful.[53] Thus, using citation counts, even with control variables, may not accurately measure the quality of the cited work.

Of course, the citation studies recognize this potential objection and some of the studies even acknowledge that taking account of citation treatment would "refine" the analysis.[54] In general, however, the authors of citation-count studies typically argue that it is not necessary to distinguish between positive and negative citations.[55] This perspective is common in citation analysis of scholarly quality, where the argument is that "an article engendering hundreds of critical comments would undoubtedly be an extremely important, albeit controversial, contribution."[56] But this assumption, although appropriate in studies of scholars, is often extended by the citation-count literature to the judicial context, where it may not belong. The argument is that negative citations, like positive citations, also reflect judicial influence, because unpersuasive decisions, at least those outside the circuit, will be simply ignored.[57] As a result, the argument concludes that the mere mention of the cited work is an indication of some degree of influence.

This argument needs to be broken down into two overlapping distinctions, both of which are important for sorting out the relationship of citations to measuring influence and quality. The first distinction is that influence is not necessarily the same as quality, and citation counts are concededly a better measure of influence than of quality, at least in the judicial context. As Landes and Posner acknowledge in another paper:

> [A] common criticism of citation analysis when it is used as an evaluative tool is inapplicable, or largely so, when it is used to study influence: that a critical citation should not be weighted as

51. *See, e.g.*, RICHARD A. POSNER, FRONTIERS OF LEGAL THEORY 422-24 (2001).

52. But as Judge Henry points out, "writing an opinion notorious for being 'wrong' might also lead to many cites." Robert Henry, *Do Judges Think? Comments on Several Papers Presented at the Duke Law Journal's Conference on Measuring Judges and Justice*, 58 DUKE L.J. 1703, 1717 (2009).

53. *See* Jake Dear & Edward W. Jessen, *"Followed Rates" and Leading State Cases, 1940-2005*, 41 U.C. DAVIS L. REV. 683, 687-88 (2007).

54. Choi & Gulati, *A Tournament of Judges?*, *supra* note 39, at 307 ("More refined methods of measuring citation counts are also possible. Those compiling citation rankings could assign a judge a positive score for favorable citations and a negative score for unfavorable citations (thereby curbing the incentive to take extreme positions in their opinions).").

55. Kosma, *supra* note 35, at 338 n.16 ("Following past practice, this study includes citations that distinguish the earlier case because such citations represent an inability to *ignore* that precedent.").

56. Shapiro, *supra* note 50, at 1543.

57. Landes et al., *supra* note 7, at 273.

> heavily as a favorable one and maybe should not be counted at all or given a negative weight. When speaking of influence rather than of quality, one has no call to denigrate critical citations. Scholars rarely bother to criticize work that they do not think is or is likely to become influential. They ignore it.[58]

Thus, negative citations may actually be *positive* indications of influence, if the alternative to the negative citation is ignoring the work. When measuring quality, on the other hand, negative citations are just that – negative. This is where a line can be drawn between the Landes et al. study and the Choi and Gulati study. The Landes et al. study and several other citation studies purport to measure the "influence" of federal judges,[59] not the "quality" of their opinions. For this purpose, not distinguishing between positive and negative citations may make sense. The Choi and Gulati approach, on the other hand, purports to use citation counts to measure "opinion quality."[60] For this purpose, treating positive, negative, and neutral citations alike may overlook the most important piece of information: the citing judge's treatment of the cited opinion. The use of citation counts to assess judges, therefore, is much more properly considered a measure of *influence*, rather than of quality or reputation.[61]

The second distinction is suggested by the first – namely that negative citations, even if not a measure of quality of work, may be a measure of influence of that work. There is an important assumption underlying this argument, however, which is that the alternative to negative citation is no citation at all. As suggested by the passage above, that may be an appropriate behavioral assumption for citations in scholarship, where unpersuasive work is regularly ignored. But the citation-count studies, with one notable exception,[62] assume that the prevailing behavioral norms in scholarship – ignoring unpersuasive work – translate into judicial opinions. In reality, however, judges often do not simply ignore the arguments of the losing party in their

---

58. William M. Landes & Richard A. Posner, *The Influence of Economics on Law: A Quantitative Study*, 36 J.L. & ECON. 385, 389-90 (1993).

59. *See* Landes et al., *supra* note 7, at 271 (explaining the purpose of their study was to "estimate empirically the influence of individual judges"); *see also* Kosma, *supra* note 35, at 333 ("This empirical study measures the influence of 99 retired Supreme Court justices . . . .").

60. Choi & Gulati, *supra* note 5, at 54 (noting that the authors use "citations as a proxy for quality" and acknowledging that some may find that "outrageous"); *see also* Choi et al., *supra* note 37, at 1321 (noting that the number of citations is used "as a proxy for the intrinsic quality of the reasoning in the opinion"). Note that Choi, Gulati, and Posner recognize the distinction between influence and quality but argue that influence and quality are highly correlated. *See* Choi et al., *supra* note 37, at 1321-22.

61. Michael E. Solimine, *Judicial Stratification and the Reputations of the United States Courts of Appeals*, 32 FLA. ST. U. L. REV. 1331, 1339 (2005).

62. *See* Dear & Jessen, *supra* note 53.

opinions, they engage those arguments, distinguishing or rejecting them if necessary. Thus, the alternative to citation is not to ignore the previous decision, but rather to cite it negatively. As a result, the mere fact that a judge cites a prior decision is not necessarily an indication of influence and certainly not an indication of quality but rather an indication of the arguments put forth by the litigants.

2. Inside- Versus Outside- Circuit Citations

The discussion of influence and quality in citation studies leads to the second drawback of those studies – their emphasis on "outside-circuit" citations as measures of influence or quality. As explained above, negative citations may reflect influence when the alternative to a negative citation is no citation at all. But even the citation studies recognize that in judicial decisions, the doctrine of binding precedent might require a citation – even a positive citation – to an unpersuasive or poorly reasoned opinion.[63] In contrast, the assumption is that a decision in another circuit will simply be ignored if it is not well-reasoned and influential. Thus, the authors of the citation studies argue that outside-circuit citations are driven by persuasion and inside-circuit citations are driven by precedent,[64] leading those authors to focus primarily on outside-circuit citations and to ignore inside-circuit citations.

The emphasis on outside-circuit citations, however, rests on problematic assumptions about how judges deal with unpersuasive decisions, both inside and outside their circuits. Although most people agree that judges cite well-reasoned, persuasive opinions positively, what do judges do with unpersuasive opinions? The citation studies assume that judges ignore those decisions if they are outside the circuit, unless they are so influential that they are impossible to ignore. This assumption leads Landes et al. to argue that "[c]ritical citations, in particular to opinions outside the citing circuit, are also a gauge of influence since it is easier to ignore an unimportant decision than to spell out reasons for not following it."[65] Thus, citations from outside the circuit are assumed to be positive because they are only persuasive.

The citation studies assume that unpersuasive decisions *inside* the circuit, on the other hand, will be cited and cited positively. The argument is that citations from the same circuit are uninformative about influence or quality because those citations are uniformly positive and compelled by precedent.[66] But this mechanical view of precedent ignores the fact that judges have a great deal of flexibility to avoid poorly reasoned opinions in their own circuits. Judges may well be bound by those decisions as

63. *See* Landes et al., *supra* note 7, at 272-73.

64. *Id.* at 285-86.

65. *Id.* at 273.

66. The Dear and Jessen study, which recognizes the importance of the distinction between positive and negative citations, still focuses on "out-of-state" citations. *See* Dear & Jessen, *supra* note 53, at 692 n.16.

precedent, but judges can often distinguish the binding precedent if they find it unpersuasive. Judges' options are not limited to following, rejecting, or ignoring a decision; judges can, in many cases, distinguish an unpersuasive decision and avoid its effect. Thus, the extent to which precedents are distinguished rather than followed is a negative indication of the persuasiveness of the precedent.

With respect to citations from other circuits, the mere fact of citation is not always an indication of quality, as suggested above, or even influence. This is because when there is no law directly on point – a likely situation in citation studies based on published opinions – the litigants will usually cite authority from other circuits. In such cases, the mere fact of citation to out-of-circuit decisions simply does not communicate anything about those decisions, other than the fact that they favored and were cited by one of the litigants. This is because, "[u]nlike scholars, courts often are not free simply to ignore authority that is, for example, expressly relied upon in a party's brief, but which the court finds unpersuasive. Instead, a court often will cite that authority and in the process criticize or at least distinguish it."[67]

Generally, when a judge cites persuasive authority from another circuit, he or she will need to respond to the decision from the other circuit, which means following, distinguishing, or criticizing the other circuit's decision. In any one of these treatments, the other circuit's decision will be cited.

There are two broad lessons from this discussion. First, inside-circuit citations still convey meaningful information if the study distinguishes between positive (e.g., following) and negative (e.g., criticizing or distinguishing). Second, outside-circuit citations do not necessarily convey meaningful information about influence or persuasiveness but rather may simply reflect the generality of a decision's reasoning or the capaciousness of its dicta. As a result, at least when measuring quality rather than influence, it makes sense to distinguish between positive and negative citations and to use inside- and outside-circuit citations.

3. Opinion Authorship Versus Panel Membership

The third problem with citation-count studies is the reliance on opinion authorship rather than panel membership to assess influence or quality.[68] These studies evaluate judges based on the citations to the opinions they write rather than to the panels on which they participate, because authoring an opinion is more closely tied to the individual judge than is serving on a panel.[69] The advantage of the opinion authorship strategy is that fewer observations

67. *Id.* at 688 n.7.

68. This problem does not affect the Cross and Lindquist reversal rate study to the same degree, as the authors of that study link judges to opinions they join, not merely to those they write. Cross & Lindquist, *supra* note 18, at 1410.

69. *See, e.g.*, Choi & Gulati, *supra* note 5, at 40-41.

need to be collected because the random variation of other factors is attenuated when judges are linked to decisions through their opinion authorship. But as long as there is sufficient data and panel membership is determined randomly, linking judges to decisions using panel membership rather than opinion authorship allows measurement of much more than opinion authorship, as discussed below.

The advantages of using panel membership rather than opinion authorship are that the analysis can (1) mitigate the effects of selection bias in opinion assignment and (2) capture collegial factors that should enter into a measure of good judging. Because opinion authorship is not randomly assigned, the measure of quality may be biased by selection effects. Opinion assignment might affect citation counts,[70] especially when opinion assignment is combined with selective publication of opinions.[71] On the other hand, using panels, rather than authored opinions, eliminates the concern of self-selection, as judges are assigned randomly to panels.[72] Moreover, this approach minimizes concerns that critics of judge rankings systems have had with gamesmanship of the rankings,[73] because it would be much more difficult to game randomly assigned panels than deliberately assigned opinions on the panels. Although judges might be able to selectively choose opinions to write or manipulate those opinions to maximize citations, it would be very hard to do that as one member of a three-judge panel.

The second advantage to linking performance to panel membership rather than opinion authorship is that panel membership might capture the intangible contributions that high-quality judges make other than writing opinions. Presumably, judges who serve on panels make some positive contribution in the cases in which they participate, even if they do not write the majority opinion. Participation in deliberations, comments on the opinion, discussion of the rationale, and even the actual vote on the disposition could all affect the persuasiveness of the resulting decision. Judge Jay Bybee (writing with Professor Thomas Miles) explains how other members of the panel contribute to the opinion ultimately produced:

---

70. For example, "[a] landmark opinion in a relatively narrow area of law . . . may be cited less than a mediocre opinion in a frequently litigated substantive area simply because there are fewer opinions generated in that area." Baker et al., *supra* note 5, at 1660.

71. *See* James J. Brudney, *Foreseeing Greatness? Measurable Performance Criteria and the Selection of Supreme Court Justices*, 32 FLA. ST. U. L. REV. 1015, 1020 (2005).

72. JONATHAN MATTHEW COHEN, INSIDE APPELLATE COURTS: THE IMPACT OF COURT ORGANIZATION ON JUDICIAL DECISION MAKING IN THE UNITED STATES COURTS OF APPEALS 72 (2002). The circuits do get different types of cases, and this could affect how judges within a circuit are viewed relative to judges in another circuit.

73. *See infra* Part V.A.

> A judge may contribute mightily to the quality of an opinion even if she is not its author. A thoughtful judge may ask penetrating questions from the bench that help shape the views of the other members of the panel. In conference discussions or in commenting on a colleague's draft opinion, a judge may influence an opinion's analysis.[74]

Similarly, Judge Harry Edwards of the United States Court of Appeals for the District of Columbia Circuit argues that, "[d]uring the course of judicial deliberations, judges more often than not persuade one another until a consensus is reached."[75] Thus, the ultimate opinion, which represents the "consensus," is the joint product of the entire panel, not merely of the opinion's author. As Choi and Gulati pointed out in their original "tournament" article, there is an important "team" aspect to the appellate panel that should be measured.[76]

Yet these contributions are not measured by techniques that focus exclusively on ranking judges by opinions they have authored. If judges are ranked based on the decisions in which they participate in addition to the opinions they author, the rankings reward rather than discourage fruitful collaboration, as some say Choi and Gulati's tournament does.[77] Indeed, the failure to capture deliberation, collegiality, and consensus-building is one of the prime criticisms of those who criticize empirical analysis of judging gen-

---

74. Jay S. Bybee & Thomas J. Miles, *Judging the Tournament*, 32 FLA. ST. U. L. REV. 1055, 1066 (2005).

75. Harry T. Edwards & Michael A. Livermore, *Pitfalls of Empirical Studies That Attempt to Understand the Factors Affecting Appellate Decisionmaking*, 58 DUKE L.J. 1895, 1964 (2009).

76. *See* Choi & Gulati, *A Tournament of Judges?*, *supra* note 39, at 308 ("A more complicated issue arises if one views an opinion as a team product. Circuit court decisions are generally rendered in groups of three. The question is how to allocate credit (or blame) for the final product. One judge writes the opinion, but the theory is that she writes the opinion in consultation with the other two judges. Our tournament gives only the writing judge credit for an opinion. Citation counts therefore represent but a noisy indicator of judicial quality. Indeed, in certain circumstances, the credit for a superb (or not-so-superb) opinion should be allocated among the three."). The authors of that study thought this problem might be "unsolvable," as they argued it would require an outsider to evaluate the relative contribution of the three. *Id.* As discussed later, the approach in this Article attempts to disentangle these contributions without an evaluator. *See infra* Part IV.C.

77. Choi & Gulati, *A Tournament of Judges?*, *supra* note 39, at 309 ("A critic might then ask whether the tournament presents the danger of discouraging collaboration. Will the two nonwriting judges refuse to work with or worse, attempt to undermine the writing judge in an effort to boost their own relative rankings?"). Choi and Gulati conclude that they "expect the competition to be healthy" but concede that the objection "is a fair one." *Id.*

erally, not merely judge ranking studies in particular.[78] Linking judges to citations based on panel membership may help to capture some of the intangibles of collegiality that quantitative studies are often criticized for ignoring.

4. Productivity and Quality

The final criticism of the citation-count studies centers on their preference for quantity over quality in opinion writing. The citation-count studies implicitly favor judges who produce more published opinions because more published opinions produce more citations. Choi and Gulati go further, explicitly incorporating "productivity" into their rankings, because they believe productivity is one characteristic of a promising Supreme Court justice.[79] But by rewarding separate opinion writing in both the productivity and independence ranking categories, Choi and Gulati's method may actually measure negative traits of judges.[80] Moreover, although productivity is perhaps an important trait for judges on the courts of appeals, it does not seem particularly relevant for Supreme Court justices.[81] Thus, a better approach would be one that avoids this "volume" measure of performance entirely, which is exactly the approach outlined in the next section.

C. Theory of This Article

The considerations discussed above suggest that the citation-count studies probably reflect a mix of judicial productivity, influence, aggressiveness, and possibly creativity or originality, but not necessarily "quality" in the usual sense.[82] As Cross and Lindquist wrote, "The quality judges on the Choi and Gulati measure appear to be fairly aggressive in their decisionmaking, provoking more frequent Supreme Court review. They are relatively successful in achieving higher numbers of affirmances, but they also suffer more losses than the average judge."[83] Or, as Judge Richard Posner put it, the evaluative criteria in the citation studies "implicitly treat judicial creativity as the only, or at least the most important, attribute of a circuit judge."[84] Although

78. *See, e.g.*, Edwards & Livermore, *supra* note 75, at 1917-18, 1951-52, 1963-66.

79. *See* Choi & Gulati, *supra* note 5, at 47.

80. Marshall, *supra* note 9, at 128-29 (arguing that the independence measure "rewards dogmatism and obstreperousness more than the type of independence we most value in our jurists"). Justices on the Supreme Court will write approximately the same number of majority opinions, and the separate opinions on the Court are not always helpful. *Id.* at 133-34.

81. *See* Bybee & Miles, *supra* note 74, at 1073-74; Cross & Lindquist, *supra* note 18, at 1389.

82. *See supra* Part II.B.

83. Cross & Lindquist, *supra* note 18, at 1413.

84. POSNER, *supra* note 15, at 148.

these may be positive traits of appellate judges,[85] they do not directly measure what most people think of as judicial quality, as opposed to scholarly quality or judicial or scholarly influence. This Article aims to develop a measure specifically directed to judicial quality.

The first step in devising a measurement strategy for judicial quality is to articulate a clear theoretical mechanism that links the unobservable features of judicial quality to observable judicial outputs. As described above, many judicial traits, such as productivity, influence, aggressiveness, creativity, originality, and quality contribute to the number of citations to the decisions of a particular judge.[86] To help isolate the effect of quality, this Article focuses not on the number of citations but on the relative proportion of positive to negative citations to a judge's decisions. The theory underlying this measurement strategy is simple: the quality of an opinion's reasoning matters to judges in their citation practices. This theory assumes that judges tend to cite high-quality opinions more positively than low-quality opinions. As a result, in the absence of binding precedent, judges will be more likely to "follow" the reasoning of high-quality opinions and more likely to criticize or distinguish the reasoning of low-quality opinions.

But what about cases in which judges face binding precedents, such as those within the same circuit? In such cases, judges generally are not free to simply "criticize" or disregard the precedent, as the very idea of binding precedent is that it must be followed or distinguished.[87] But the power to distinguish a precedent leaves judges with considerable freedom to avoid decisions with which they do not agree. Judges may distinguish a prior case because the rule of the precedent does not apply to the facts of the present case, but judges may also distinguish a precedent because the precedent was not well-reasoned, although purporting to identify facts that distinguish the prior case. It is always possible to distinguish a precedent (although perhaps at the cost of making the distinguishing case itself less persuasive), so even in the presence of binding authority the proportion of cases "following" versus

---

85. They may not be, however. Professor Solum argues that "originality" is characteristic of bad judges, and that originality also stimulates citations, creating a problem for citation-count studies:

> Good judges are clever in using the resources within existing law to solve the legal problems that come before them. The very best judges are experts at avoiding originality. And the very worst judges may be the most original. Very bad judges may use the cases that come before them as vehicles for changing the law, transforming the rules laid down into the rules that they prefer. This kind of results-oriented or legislative judging may produce many original propositions of law and hence a high citation rate, but this is a measure of judicial vice and not judicial virtue.

Solum, *supra* note 43, at 1393.

86. *See supra* Part II.B.

87. *See* JOSEPH RAZ, THE AUTHORITY OF LAW: ESSAYS ON LAW AND MORALITY 185 (1979).

"distinguishing" a precedent will contain important information about the quality of the precedent.

There is a third reason a judge might seek to distinguish or otherwise avoid a prior decision, even if that decision is well-reasoned or precedential. The judge might simply have a preference for an outcome different from the one the prior decision would dictate. As a result, ideological differences between judges may account for some negative citations, because judges who are ideologically extreme may tend to be cited negatively, just as judges whose opinions are of lower quality will tend to be cited more negatively. Ideological behavior is often perceived as a negative characteristic in a judge,[88] and even if perceived neutrally, the more extreme the judge the less likely others will agree with him or her. Thus, both ideology and quality will probably affect the mix of positive versus negative citations in appellate courts.

The fact that both ideology and quality might affect citations, however, does not mean we cannot disentangle the two effects. Indeed, exploring the implications of ideology in judicial rankings will be the focus of Part V.D.1. Rather than attempting to make assumptions that would disaggregate the two effects, this Article presents multiple perspectives on the data and allows readers to draw their own conclusions.

III. DATA AND METHODS

The data for this study consist of 311,931 citations among published federal appellate court cases between 1960 and 2008.[89] The data include a pool of 120,906 cases that are cited by other cases (the "cited cases") and a pool of 117,280 cases that cite other cases (the "citing cases").[90] Many of the cited cases are also citing cases, so that the total number of unique cases is 170,786. The data were collected using LexisNexis's Shepard's Citations service in mid-2008, so the data does not include any cited cases or citing cases decided after that time.

The distinctive feature of this dataset is that each of the citations is coded as "positive" or "negative" according to the Shepard's Citations treatment code assigned to the citation. These treatment codes are assigned by staff attorneys who read the cases and are designed to indicate the preceden-

88. *See, e.g.*, TERRI JENNINGS PERETTI, IN DEFENSE OF A POLITICAL COURT 87, 89-90 (1999) (noting the common perception of ideological behavior as negative and arguing against that normative implication for Supreme Court justices).

89. Specifically, the cases are those in volume 273 or later of the *Federal Reporter, Second Series* and prior to volume 540 of the *Federal Reporter, Third Series*.

90. The cases only include those with a citation in the *Federal Reporter*, which generally indicates a published opinion. Excluded are opinions in which multiple cases share the same citation (i.e., when they begin on the same page of the *Federal Reporter*).

tial value of the cited cases.[91] The majority of Shepard's Citations specify no treatment other than "citing," which does not indicate any positive or negative relationship between the cited case and the citing case. But a substantial percentage of citations include a treatment code, such as "following," "distinguishing," "criticizing," "limiting," "overruling," and so forth. These treatment codes reflect the relationship between the cited case and the citing case and provide the key data for this study. Scholars have conducted extensive tests on the treatment codes in Shepard's data and found that the treatment coding is generally quite valid and reliable.[92]

The dependent variable in this analysis is a dichotomously coded outcome indicating the treatment of the citations – i.e., "positive" or "negative." Consistent with other empirical work using Shepard's Citations, "positive" codes include the Shepard's treatment of "following," and negative codes include the Shepard's treatments of "distinguishing," "criticizing," "questioning," "overruling," and "limiting."[93] Although the "distinguishing" code is not obviously negative in the same way as "criticizing," as discussed in Part II.B above, the citing case is negative in the sense that the citing case seeks to avoid applying the reasoning of the cited case. Of course, the "distinguishing" code may genuinely indicate the citing case's fact pattern is not within the holding of the cited case, but those factual distinctions should occur at random and only make the measures noisy, not biased. Table I below presents the distribution of each treatment category in the data.

Table I. Treatment Codes of Cited Cases	
Citation Treatment	*Number of Citations*
Followed	197,111
Distinguished	86,021
Questioned	10,094
Limited	308
Criticized	17,600
Overruled	794

The parameters of interest are estimated using a linear probability model[94] with a binary dependent variable regressed on a large number of indepen-

91. James F. Spriggs, II & Thomas G. Hansford, *Measuring Legal Change: The Reliability and Validity of* Shepard's Citations, 53 POL. RES. Q. 327, 330 (2000).

92. *See id.*

93. THOMAS G. HANSFORD & JAMES F. SPRIGGS II, THE POLITICS OF PRECEDENT ON THE U.S. SUPREME COURT 58 (2006).

94. Although the dependent variable is binary, the linear probability model is used for ease of interpretation of the judge identifier coefficients. Because this study does not conduct hypothesis tests of the coefficients or attempt to predict probabilities, the well-known limitations of the linear probability model with binary dependent variables are outweighed by the ease of interpretation of the coefficients for this study.

dent variables.[95] The dependent variable takes the value 1 if the citation is "positive" and 0 if the citation is "negative," as described above. This means that the linear probability model estimates relationships between the independent (explanatory) variables to whether the case is cited positively or negatively.

The independent variables of interest are indicator variables for each of the 466 judges included in the data.[96] The judge indicator variables (one for each judge) take the value of 1 if the judge served on the panel of the cited case and 0 if otherwise.[97] If a judge dissented in the cited case, the judge is treated as having been removed from the panel, meaning that the variable takes the value 0.[98] The coefficient on the judge's indicator variable, therefore, may be interpreted as the contribution that judge makes toward the case being positively (rather than negatively) cited. If a judge's presence on the panel of the cited case is associated with the case being cited more positively, the judge will have a positive coefficient. If a judge's presence on the panel of the cited case is associated with the case being cited more negatively, the judge will have a negative coefficient. These coefficients are the main quantities of interest used to estimate judicial quality. Larger positive coefficients may be interpreted as indications of higher quality and smaller (or negative) coefficients may be interpreted as indications of lower quality.

The model also includes three principal types of control variables. The first control variables are the volume numbers of the Federal Reporter for the citing case and the cited case, which serve as proxies for a time variable.[99] It is well known that the number of citations to a precedent depreciates over time,[100] and this pattern translates into more negative citations as well – older

95. The size of the data required the use of a bounded memory regression package in the R Statistical Programming Language called biglm.

96. Although 466 judges were contained in the data and therefore used in estimating the model, only 383 judges, with sufficiently small standard errors, are presented in the tables below. *See infra* Part IV.

97. The identity of the judges in each case was obtained by using the "JUDGES( )" segment search in LexisNexis, using the judge's last name and restricted to the relevant circuit. In cases where two or more judges shared the same name, the cases were distinguished by date where possible and by first names where the judges with the same name served at the same time.

98. This is because if a judge dissents from a high-quality case, he or she should not receive an increase in the quality estimate because of that case, and if a judge dissents from a low-quality case, he or she should not receive a reduction in the quality estimate because of that case. The identity of judges dissenting from opinions was obtained by using the "DISSENTBY( )" segment search in LexisNexis and disambiguating shared surnames.

99. The Second Series of the *Federal Reporter* is numbered consecutively from 1 to 999, and the Third Series restarts at 1, so 999 was added to Third Series volume numbers to keep them consecutive.

100. William M. Landes & Richard A. Posner, *Legal Precedent: A Theoretical and Empirical Analysis*, 19 J.L. & ECON. 249, 259 (1976).

cases are cited more negatively than recent cases. Thus, to make the measures comparable over time, a control variable for the volume number of the cited case is included. Moreover, older cases also cite other cases more negatively, even holding constant the date of the cited case.[101] This effect is not as strong as the first but also requires a control variable. There is an approximately linear trend in both relationships over time, so a variable is included for the volume numbers of the citing case and the cited case to control for time.[102]

The second set of variables controls for the effects of inside-circuit versus outside-circuit citations. As one might expect, judges cite cases within their own circuits much more frequently and favorably than cases outside their circuits.[103] Whether one attributes this effect to the constraining force of precedent or to the threat of en banc review, the effect is a considerable one, as the analysis in Part IV demonstrates. As a result, without control variables, if some judges tend to be cited more by their own circuit than others, those judges would appear to be of higher quality than the others, when in fact the judges receiving more outside-circuit citations might be higher quality.[104] The citation-count studies dealt with this complication by focusing primarily on outside-circuit citations. But using only outside-circuit citations means leaving out half the data, and it may have other disadvantages as well.[105]

Rather than focusing on outside-circuit citations such as those analyzed in previous studies, the approach in this Article controls for the inside-circuit effect, which also allows estimation of a separate inside-circuit effect for each circuit. The control variables therefore include twelve indicator variables, one for each circuit (the Federal Circuit is omitted because its specialized docket would make comparisons to judges in other circuits unreliable). These

---

101. The cause of this phenomenon is unclear. One possibility is changes in Shepard's coding standards over the years, but it seems unlikely that coding standards would change year after year at approximately the same rate. Another possibility is that the increased use of non-published opinions increased the quality of published opinions. Whether this is an artifact of the coding conventions used by Shepard's Citations or whether this reflects changes in the legal environment since the 1960s must be left to future research.

102. To check for non-linearities, an alternative specification with twenty-six dummy variables for different time periods was run, and the correlation between the coefficients for the two models was approximately .957, suggesting the dummy variable specification would not improve much over the linear time trend. It is worth noting that these control variables help to maintain intertemporal comparability but do impose the constraint that judge quality does not improve or decrease systematically over time, making it impossible to say whether the judiciary gets better or worse over time. This limitation may not cause significant problems, as most practical applications involve comparing the performance of contemporaneous judges, and it may not be meaningful to compare today's judges with those of the 1960s anyway.

103. *See infra* Part IV.A tbl.IV.

104. Indeed, this is the assumption that the citation-count studies make.

105. *See supra* Part II.B.2.

variables take the value 1 if the cited case and the citing case are in the same circuit and 0 otherwise. A thirteenth control variable is included for Eleventh Circuit cases that cite Fifth Circuit cases, as some Fifth Circuit opinions have precedential value in the Eleventh Circuit.[106] These variables control for in-circuit citation and also estimate the in-circuit citation effect – reflecting, in part, the constraining force of precedent.

A final set of variables attempts to capture ideological differences between the citing panel and the cited panel, using the political parties of appointing presidents as proxies for ideology. A panel with a majority of Democrat-appointed judges is coded as Democratic, and a panel with a majority of Republican-appointed judges is coded as Republican.[107] Using the political party of the appointing president as a proxy for ideology is as controversial as it is standard.[108] Of course, this variable will not capture all or perhaps even most ideological differences between panels, but the variable does reveal whether ideology, as measured by political party of the appointing president, affects whether the judge cites another case positively or negatively. As discussed below, these variables also help answer the question of the extent to which ideology and precedent affect inside-circuit and outside-circuit citations.[109]

IV. RESULTS AND INTERPRETATION

A. A Ranking of Federal Appellate Judges Since 1960

The rankings are presented in Table II for the full dataset (column 4) and five subsets of the dataset (columns 5-9), discussion of which is deferred to Part IV.B below. The judges are ranked in order from the most positively cited judges to the most negatively cited judges according to column 4 (results for the full dataset). Column 1 lists the judges' names, column 2 lists

106. The Eleventh Circuit was carved out of the Fifth Circuit in 1981. Fifth Circuit Court of Appeals Reorganization Act of 1980, Pub. L. 96-452, 94 Stat. 1995. In *Bonner v. City of Prichard*, the new Eleventh Circuit held that decisions of the Fifth Circuit handed down prior to September 30, 1981 would have binding authority in the Eleventh Circuit. 661 F.2d 1206 (11th Cir. 1981).

107. Some panels only have two judges, either because a judge from outside the circuit is sitting by designation or because one judge died or exercised the recusal option. In these cases, if the panel is comprised of one Democrat and one Republican, the data is treated as missing for this variable.

108. *See* Stephen J. Choi & G. Mitu Gulati, *Bias in Judicial Citations: A Window into the Behavior of Judges?*, 37 J. LEGAL STUD. 87, 94 (2008). The use of presidential party as a proxy for ideology has been widely criticized by scholars and by judges themselves. *See, e.g.*, Edwards & Livermore, *supra* note 75, at 1909 n.21, 1918-22. However, the practice is relatively standard. *See* CASS R. SUNSTEIN ET AL., ARE JUDGES POLITICAL?: AN EMPIRICAL ANALYSIS OF THE FEDERAL JUDICIARY 8-9 (2006).

109. *See infra* Part IV.C.

the judges' circuits, and column 3 lists the judges' "scores," which are the coefficients on their indicator variables in the full dataset regression. Larger scores indicate a judge is more positively cited and smaller scores (including negative scores) indicate a judge is more negatively cited. Only judges whose coefficients had standard errors of .024 or less in the full dataset are presented, which eliminates some judges from the ranking if their positions are more uncertain because of a small number of observations.[110] It should be noted that because of the number of parameters estimated, most of the judges' scores could vary considerably. In drawing inferences about individual judges, only those consistently toward the top or the bottom of the rankings can be considered reliable.

Table II. Rankings of Federal Appellate Judges Since 1960

Columns

1	2	3	4	5	6	7	8	9
Judge Name	Circuit	Score: All Panels	Ranking: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: By Democrats	Ranking: By Republicans
Riley	8	0.083	1	10	17	4	48	11
Selya	1	0.082	2	11	7	29	10	10
Straub	2	0.076	3	2	109	344	43	3
Smith	8	0.068	4	23	45	145	8	20
Ainsworth	5	0.068	5	20	33	46	12	36
Breyer	1	0.068	6	13	57	39	6	33
Motz	4	0.068	7	27	99	119	119	14
Carter	9	0.066	8	34	12	19	29	22
GarzaE	5	0.065	9	133	69	23	13	15
Wallace	9	0.064	10	60	22	176	92	5
Wilkins	4	0.064	11	30	41	31	132	17
Kennedy	9	0.063	12	4	166	98	140	9
Starr	12*	0.06	13	49	78	89	71	2
Ambro	3	0.06	14	36	38	157	27	37
Pickett	10	0.06	15	24	8	81	212	7
Thornberry	5	0.06	16	59	108	210	39	31
Sotomayor	2	0.059	17	1	29	9	31	25
Robb	12*	0.059	18	5	28	1	291	6
King	5	0.059	19	22	19	12	17	50
Tamm	12*	0.058	20	7	50	7	42	52

110. The value of .024 was chosen because this was the smallest value that would include Judges (later Justices) Burger and Blackmun, who are discussed *infra* Part V.A.

McKee	3	0.058	21	73	34	88	82	44
Pell	7	0.057	22	65	137	227	68	18
Skopil	9	0.057	23	120	10	24	14	67
Henderson	12*	0.057	24	17	56	146	162	16
Coffey	7	0.056	25	26	9	22	46	38
Leval	2	0.056	26	53	114	168	7	65
Johnson	11	0.055	27	61	112	107	30	95
Timbers	2	0.055	28	89	176	152	5	118
Weis	3	0.054	29	47	54	58	23	134
Russell	4	0.052	30	63	64	87	38	55
Berzon	9	0.052	31	19	42	100	20	147
Friendly	2	0.052	32	57	101	91	26	51
Phillips	4	0.052	33	109	189	251	83	21
Marcus	11	0.052	34	327	149	207	287	8
Ebel	10	0.052	35	29	43	42	146	27
Ripple	7	0.051	36	145	11	41	84	57
Howard	1	0.051	37	3	13	15	1	160
Hansen	8	0.051	38	80	141	167	70	48
Tjoflat	5	0.05	39	155	261	351	19	186
WoodD	7	0.05	40	94	74	132	93	56
O'Scannlain	9	0.05	41	9	195	68	115	42
Baldock	10	0.05	42	128	81	96	214	23
Nygaard	3	0.049	43	25	3	8	72	47
Flaum	7	0.048	44	135	30	84	222	45
Lynch	1	0.048	45	48	91	90	65	112
Campbell	1	0.048	46	35	63	38	76	128
Rogers	12*	0.048	47	97	53	13	16	100
MacKinnon	12*	0.047	48	143	88	120	95	90
Wilkey	12*	0.047	49	37	20	55	55	114
Graber	9	0.047	50	21	75	35	172	12
Murphy	8	0.047	51	102	134	280	166	61
Aldisert	3	0.047	52	75	52	83	204	32
Altimari	2	0.046	53	70	72	137	109	97
Bownes	1	0.046	54	198	49	181	57	155
Scirica	3	0.046	55	31	132	136	9	143
Walker	2	0.046	56	54	44	60	176	102
Trott	9	0.045	57	46	238	305	225	19
Ross	8	0.045	58	169	103	133	122	49
Blackmun	8	0.045	59	312	5	187	365	1
Medina	2	0.045	60	273	217	127	61	62

Table II. Rankings of Federal Appellate Judges Since 1960 (cont)

1	2	3	4	5	6	7	8	9
Judge Name	Circuit	Score: All Panels	Ranking: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: By Democrats	Ranking: By Republicans
Katzmann	2	0.045	61	151	110	62	99	149
GibsonJ	8	0.045	62	125	136	162	249	24
Matthes	8	0.045	63	295	4	34	128	43
Barnes	9	0.044	64	265	241	281	123	69
Robinson	12*	0.044	65	112	146	249	88	101
Burger	12*	0.044	66	225	299	212	173	4
Williams	7	0.044	67	121	92	95	279	30
Stephenson	8	0.044	68	84	202	185	106	54
Newman	2	0.044	69	81	159	101	90	75
McKeown	9	0.044	70	101	24	77	117	41
Cabranes	2	0.044	71	182	16	52	87	120
Sprecher	7	0.043	72	55	139	47	4	131
Williams	12*	0.043	73	40	180	121	192	99
Kearse	2	0.043	74	150	47	108	175	80
Sloviter	3	0.042	75	28	158	65	86	89
Garland	12*	0.042	76	197	2	3	170	74
Adams	3	0.042	77	241	144	253	28	191
WoodH	7	0.042	78	98	80	66	85	126
Roth	3	0.042	79	51	201	197	32	119
Alito	3	0.042	80	64	151	92	261	34
Carnes	11	0.041	81	154	130	250	98	107
Milburn	6	0.041	82	106	76	174	97	85
Tacha	10	0.04	83	147	26	40	194	78
Gilman	6	0.04	84	15	170	50	188	13
Celebrezze	6	0.04	85	117	48	149	21	205
Bowman	8	0.04	86	140	82	71	250	40
Cudahy	7	0.04	87	122	113	49	158	124
Moore[7]	10	0.04	88	131	332	282	94	136
Barksdale	5	0.04	89	58	89	21	96	86
Miner	2	0.04	90	105	153	233	15	206
Lucero	10	0.04	91	42	73	14	118	103
Rubin	5	0.039	92	76	187	139	144	108
Birch	11	0.039	93	32	211	186	40	154

Hollo-way	10	0.039	94	110	169	115	145	83
Clark	11	0.039	95	176	118	170	89	84
Kanne	7	0.039	96	253	21	159	227	76
Brorby	10	0.039	97	86	61	54	139	96
Torruel-la	1	0.039	98	99	65	184	240	87
Wil-liams	5	0.039	99	104	163	126	124	91
Easter-brook	7	0.038	100	152	100	118	309	70
Farris	9	0.038	101	96	107	74	235	73
Ryan	6	0.038	102	52	221	269	329	26
Feinberg	2	0.037	103	108	102	129	186	137
Parker	5	0.037	104	41	119	16	111	151
Kennedy	6	0.037	105	190	115	172	181	39
Buckley	12*	0.037	106	161	32	122	161	77
Fagg	8	0.037	107	66	230	93	36	130
Hastings	7	0.037	108	219	258	175	187	46
Cyr	1	0.037	109	163	15	76	231	117
Roney	5	0.037	110	227	1	2	112	133
Meskill	2	0.036	111	111	111	110	44	175
Henley	8	0.036	112	141	190	117	49	161
Kravitch	11	0.036	113	79	210	204	74	196
Roney	11	0.036	114	67	240	72	236	72
Bright	8	0.035	115	45	346	153	54	221
Luttig	4	0.035	116	279	192	260	347	28
Woll-man	8	0.035	117	149	68	123	201	68
Dennis	5	0.035	118	8	66	64	3	226
Wald	12*	0.035	119	177	116	111	66	181
Wilson	11	0.035	120	95	215	112	195	104
Murphy	10	0.035	121	85	71	36	78	125
Keith	6	0.035	122	184	275	274	137	105
Hull	11	0.035	123	71	51	230	147	71
Michael	4	0.035	124	77	243	192	125	113
Rymer	9	0.034	125	16	280	223	282	220
Staley	3	0.034	126	138	157	109	2	339
Tate	5	0.034	127	12	246	28	101	148
Bye	8	0.034	128	14	228	82	11	195
Cox	11	0.034	129	144	311	346	257	110
Hall	4	0.034	130	116	281	231	81	211
Clark	5	0.034	131	92	148	70	298	58
Boudin	1	0.034	132	211	160	75	300	59
Hamlin	9	0.034	133	160	121	158	107	35
Manion	7	0.034	134	172	98	114	217	129

Table II. Rankings of Federal Appellate Judges Since 1960 (cont)

1	2	3	4	5	6	7	8	9
Judge Name	Circuit	Score: All Panels	Ranking: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: By Democrats	Ranking: By Republicans
Wilkinson	4	0.033	135	159	183	196	330	88
Winter	2	0.033	136	164	161	147	73	217
Moore	6	0.033	137	119	152	203	41	266
Duhe	5	0.033	138	205	198	271	62	150
HillJ	5	0.033	139	291	18	150	60	254
Evans	7	0.033	140	44	25	17	47	216
Hays	2	0.033	141	18	97	6	33	305
Barrett	10	0.033	142	178	167	320	80	185
Anderson	11	0.033	143	88	120	163	155	173
McLaughlin	2	0.032	144	212	85	113	91	253
Bell	5	0.032	145	303	23	213	77	194
Kelly	10	0.032	146	78	147	18	237	106
DeMoss	5	0.032	147	50	172	30	288	115
GibsonF	8	0.032	148	269	216	242	130	180
Wiggins	9	0.032	149	209	206	294	18	311
Gewin	5	0.032	150	323	87	155	102	94
Hill	11	0.032	151	186	173	182	142	123
Hutchinson	3	0.032	152	103	105	94	306	98
Guy	6	0.032	153	114	254	296	135	172
Bauer	7	0.032	154	185	39	106	152	165
Briscoe	10	0.031	155	193	286	355	58	273
Kilkenny	9	0.03	156	374	46	343	274	132
Seitz	3	0.03	157	262	31	154	67	252
Williams	4	0.03	158	91	79	56	320	152
Hawkins	9	0.03	159	294	77	124	179	209
McKay	10	0.03	160	56	174	73	216	142
Eschbach	7	0.029	161	226	94	67	151	171
Gee	5	0.029	162	162	328	221	56	189
Anderson	10	0.029	163	246	104	143	169	140
Murnaghan	4	0.028	164	222	219	265	50	284
Godbold	11	0.028	165	83	224	135	136	92
Siler	6	0.028	166	132	218	128	325	79
McLaughlin	3	0.028	167	82	268	188	113	144

Boo-chever	9	0.028	168	210	260	218	223	109
Brei-tenstein	10	0.028	169	191	213	180	165	183
Wellford	6	0.028	170	118	363	357	63	265
Trask	9	0.028	171	343	257	364	256	153
Maho-ney	2	0.027	172	201	296	222	100	199
Oakes	2	0.027	173	171	255	313	255	167
Choy	9	0.027	174	194	138	225	248	176
Sentelle	12*	0.027	175	187	142	37	233	139
Gibbons	3	0.027	176	239	93	103	22	246
Lively	6	0.027	177	204	292	291	164	290
Rovner	7	0.027	178	167	106	48	141	158
Garth	3	0.026	179	233	154	216	281	163
HillR	5	0.026	180	33	59	5	326	218
Boggs	6	0.026	181	213	264	293	269	53
Ran-dolph	12*	0.025	182	338	70	226	24	297
Mehaffy	8	0.025	183	235	282	241	108	166
Godbold	5	0.025	184	228	197	292	114	190
Fay	11	0.025	185	290	333	332	182	164
Niemey-er	4	0.025	186	107	140	61	335	135
Bryan	4	0.024	187	136	375	374	244	198
Sneed	9	0.024	188	189	177	238	246	162
Stewart	5	0.024	189	38	62	11	323	159
Edwards	12*	0.024	190	257	124	256	211	250
Jolly	5	0.024	191	276	266	331	103	193
Logan	10	0.024	192	123	186	131	322	93
Hatchett	11	0.024	193	254	199	246	138	141
Edmond son	11	0.024	194	223	122	44	327	64
Martin	6	0.024	195	232	193	142	226	145
Ward-law	9	0.024	196	363	316	383	149	345
Hunter	3	0.024	197	170	123	134	59	215
Coffin	1	0.024	198	243	40	45	110	222
Hall	9	0.023	199	6	209	27	364	63
Tang	9	0.023	200	39	249	235	177	187
Becker	3	0.023	201	180	244	248	229	232
Cum-mings	7	0.023	202	158	233	201	185	202
Field	4	0.023	203	93	276	43	143	237
Van Ooster-hout	8	0.023	204	207	310	286	253	228

Table II. Rankings of Federal Appellate Judges Since 1960 (cont)

1	2	3	4	5	6	7	8	9
Judge Name	Circuit	Score: All Panels	Ranking: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: By Democrats	Ranking: By Republicans
Doyle	10	0.023	205	252	290	255	203	235
Lipez	1	0.022	206	358	95	99	52	256
Brown	5	0.022	207	281	165	195	202	271
Haynsworth	4	0.022	208	258	273	337	316	60
Mansfield	2	0.022	209	240	178	191	64	213
Craven	4	0.021	210	236	270	330	221	179
Morgan	5	0.021	211	175	214	237	134	321
Phillips	6	0.021	212	174	133	102	349	29
Gurfein	2	0.021	213	318	277	283	228	188
Boreman	4	0.021	214	247	242	323	232	169
Black	11	0.021	215	203	335	258	105	230
Davis	5	0.021	216	200	37	232	339	156
Gould	9	0.021	217	129	128	104	37	356
Ervin	4	0.02	218	148	191	141	163	269
Tjoflat	11	0.02	219	231	171	236	218	201
Peck	6	0.02	220	156	331	329	153	234
Wisdom	5	0.02	221	234	131	144	220	231
Widener	4	0.02	222	292	207	234	345	116
Anderson	5	0.02	223	355	27	365	171	146
Fisher	9	0.019	224	381	262	371	45	376
Hufstedler	9	0.019	225	316	347	314	79	342
Calabresi	2	0.019	226	314	156	257	51	299
NelsonT	9	0.019	227	347	222	336	278	301
Mulligan	2	0.019	228	195	60	32	35	262
Webster	8	0.018	229	329	36	105	180	312
Chapman	4	0.018	230	179	150	20	318	258
Poole	9	0.018	231	287	129	80	156	295
McEntee	1	0.018	232	153	181	217	126	208
Dyer	5	0.018	233	238	204	79	266	192
Loken	8	0.017	234	208	250	156	368	184
Lay	8	0.017	235	277	323	261	127	302
Butzner	4	0.017	236	215	208	239	270	239
Tatel	12*	0.017	237	275	279	219	286	81
ParkerB	2	0.017	238	217	35	85	243	66

GinsburgD	12*	0.017	239	300	225	334	267	236
Hamilton	4	0.016	240	206	14	25	334	177
Vance	11	0.016	241	68	269	228	133	111
Daughtrey	6	0.016	242	146	344	301	258	337
Moore	2	0.016	243	284	234	278	160	241
GarzaR	5	0.016	244	69	278	302	294	233
Contie	6	0.016	245	320	117	229	210	310
Pooler	2	0.016	246	305	326	309	206	168
GinsburgR	12*	0.016	247	74	168	10	239	229
Melloy	8	0.016	248	249	251	266	374	219
Cowen	3	0.016	249	229	58	33	157	287
Jacobs	2	0.015	250	124	127	166	328	157
Schroeder	9	0.015	251	188	305	349	174	264
Clay	6	0.015	252	181	162	183	215	291
McWilliams	10	0.015	253	299	125	51	273	270
Smith	5	0.015	254	268	239	252	361	138
JonesE	5	0.015	255	334	182	272	289	203
Van Dusen	3	0.015	256	301	196	178	184	316
Leventhal	12*	0.015	257	319	6	63	121	350
Pratt	2	0.015	258	244	135	160	312	248
Vogel	8	0.014	259	115	372	363	302	122
FletcherW	9	0.014	260	370	145	189	53	364
Garwood	5	0.014	261	307	248	264	208	245
Kaufman	2	0.014	262	272	295	317	238	306
Arnold, R.S.	8	0.014	263	62	312	215	254	224
Browning	9	0.014	264	266	317	353	333	210
Greenberg	3	0.014	265	230	155	97	193	304
Magill	8	0.014	266	286	203	169	264	212
Benavides	5	0.014	267	371	194	303	277	204
Barkett	11	0.014	268	216	304	339	224	332
Posner	7	0.013	269	251	236	240	168	278
Rosenn	3	0.013	270	311	247	326	307	182
Goldberg	5	0.013	271	298	237	304	159	336
Duniway	9	0.013	272	168	253	220	197	281

Table II. Rankings of Federal Appellate Judges Since 1960 (cont)

1	2	3	4	5	6	7	8	9
Judge Name	Circuit	Score: All Panels	Ranking: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: By Democrats	Ranking: By Republicans
Koelsch	9	0.013	273	351	327	379	358	121
Lumbard	2	0.012	274	166	293	208	251	293
Higginbotham	5	0.012	275	139	313	244	293	238
Alarcon	9	0.012	276	100	314	198	259	280
Smith	2	0.012	277	317	308	202	265	286
Krupansky	6	0.012	278	322	294	290	25	322
Stapleton	3	0.012	279	157	188	161	219	288
Vance	5	0.012	280	165	212	347	283	282
Henry	10	0.011	281	126	298	164	205	296
Wright	12*	0.011	282	308	337	333	150	353
Dubina	11	0.011	283	127	143	148	178	214
Batchelder	6	0.011	284	113	321	194	344	178
JohnsonS	5	0.011	285	220	232	177	207	251
Hug	9	0.011	286	199	226	267	247	259
Seymour	10	0.01	287	250	284	273	189	314
Weick	6	0.01	288	304	362	361	303	283
Thompson	9	0.009	289	271	291	284	341	127
Politz	5	0.009	290	267	274	254	271	244
Sprouse	4	0.009	291	337	83	130	352	227
Henderson	11	0.009	292	274	86	59	260	276
Biggs	3	0.008	293	218	223	245	198	260
Anderson	9	0.008	294	293	179	190	346	207
Suhrheinrich	6	0.008	295	310	126	276	199	247
JohnsonF	5	0.008	296	43	67	86	319	362
Cole	6	0.007	297	137	289	211	209	326
Wiener	5	0.007	298	288	252	288	317	240
Fairchild	7	0.007	299	134	353	279	116	357
Goodwin	9	0.007	300	349	330	352	295	272
Mansmann	3	0.007	301	256	184	179	183	343
Simpson	5	0.007	302	348	319	307	336	200
Pierce	2	0.006	303	297	229	209	69	308
Engel	6	0.006	304	302	272	311	290	267

Hamley	9	0.006	305	278	302	315	191	334
Paez	9	0.006	306	378	342	382	104	383
Nelson	6	0.006	307	173	303	26	366	225
Fernandez	9	0.006	308	90	350	277	304	243
Stevens	7	0.005	309	342	315	310	296	274
Van Graafeiland	2	0.005	310	270	318	206	301	324
Wright	9	0.005	311	283	231	297	262	313
Canby	9	0.005	312	183	349	318	354	170
McMillian	8	0.004	313	192	245	224	252	307
MillerW	6	0.004	314	248	371	377	351	255
Porfilio	10	0.004	315	260	175	262	337	285
Mikva	12*	0.004	316	72	235	151	275	335
Brown	6	0.004	317	202	90	78	377	242
Bazelon	12*	0.003	318	321	361	362	285	289
Pregerson	9	0.003	319	330	288	324	129	373
Leavy	9	0.003	320	87	354	69	148	359
Norris	6	0.003	321	237	220	116	314	268
Lewis	10	0.003	322	313	283	306	276	309
Ely	9	0.003	323	326	256	350	280	315
Hill	10	0.002	324	377	84	299	268	340
Ingraham	5	0.002	325	296	334	316	353	275
Reavley	5	0.001	326	142	325	243	310	319
Sack	2	0.001	327	196	373	376	131	370
Stahl	1	0.001	328	264	205	140	272	328
Bork	12*	0.001	329	353	307	342	380	82
Seth	10	0	330	309	370	370	190	292
Kerner	7	0	331	245	301	53	313	317
Murrah	10	0	332	332	227	263	360	174
Aldrich	1	0	333	221	324	214	362	263
Heaney	8	0	334	263	365	298	154	347
Tuttle	5	0	335	369	285	319	350	197
Tone	7	0	336	130	271	138	308	320
Cardamone	2	0	337	352	185	270	332	300
McGowan	12*	-0.001	338	336	164	171	120	358
Rives	5	-0.001	339	285	360	268	355	223
Coleman	5	-0.003	340	315	343	367	305	352
Rendell	3	-0.004	341	361	200	327	348	257
Tuttle	11	-0.004	342	373	340	366	213	368
Arnold, M.S.	8	-0.004	343	255	259	193	242	331

Table II. Rankings of Federal Appellate Judges Since 1960 (cont)

1	2	3	4	5	6	7	8	9
Judge Name	Circuit	Score: All Panels	Ranking: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: By Democrats	Ranking: By Republicans
Kozinski	9	-0.004	344	261	322	173	343	298
Winter	4	-0.005	345	339	300	289	196	354
Traxler	4	-0.005	346	224	96	125	75	303
Jones	6	-0.005	347	346	339	335	292	333
Higginbotham	3	-0.005	348	335	287	200	263	330
NelsonD	9	-0.005	349	359	338	378	167	369
Hastie	3	-0.005	350	259	377	360	200	325
Brunetti	9	-0.006	351	324	309	341	245	363
Norris	9	-0.006	352	242	267	205	370	261
Lewis	3	-0.007	353	280	320	199	34	371
Merrill	9	-0.009	354	350	306	321	311	344
Kalodner	3	-0.009	355	333	381	372	375	249
Merritt	6	-0.009	356	325	352	340	338	323
Beam	8	-0.01	357	344	263	322	359	327
Knoch	7	-0.01	358	306	379	359	324	346
Edwards	6	-0.011	359	367	341	300	234	367
Thomas	9	-0.011	360	360	55	57	315	279
Castle	7	-0.013	361	331	366	345	356	341
Silberman	12*	-0.015	362	214	367	247	372	338
Cecil	6	-0.015	363	340	336	348	340	329
Beezer	9	-0.015	364	282	348	275	284	360
Anderson	2	-0.016	365	356	265	308	369	318
Chambers	9	-0.017	366	357	351	295	378	349
Swygert	7	-0.017	367	328	358	325	297	366
FletcherB	9	-0.022	368	375	369	358	230	378
Fay	5	-0.025	369	368	345	356	342	374
McCree	6	-0.025	370	362	355	338	241	377
Sobeloff	4	-0.027	371	364	329	285	321	381
Waterman	2	-0.027	372	365	356	312	379	355
JonesW	5	-0.031	373	345	368	259	363	372
Noonan	9	-0.033	374	366	376	354	376	361
Duffy	7	-0.035	375	354	357	287	367	348
Ferguson	9	-0.037	376	372	374	368	357	375
Jertberg	9	-0.037	377	341	378	373	331	365
Kiley	7	-0.039	378	289	382	369	371	351

Klein-feld	9	-0.04	379	376	297	165	381	294
Rein-hardt	9	-0.043	380	380	364	375	299	380
OSulli-van	6	-0.046	381	379	380	380	383	277
Fahy	12*	-0.063	382	383	359	328	373	382
Schnack enberg	7	-0.084	383	382	383	381	382	379

** In this and all other tables, the value of 12 in column 2 indicates the D.C. Circuit.*
[1] John C. Porfilio served on the Tenth Circuit under the name John P. Moore until January 8, 1996.

The results of this analysis differ dramatically from those of prior judge ranking studies.[111] The top ten percent of the rankings (roughly ranks 1-40) contains some well-known judges, such as now-Justices Stephen Breyer, Anthony Kennedy, and Sonia Sotomayor, as well as Judges Henry Friendly and Kenneth Starr, among others. In addition, the top of the rankings contains some names that also stood out in Choi and Gulati's measure, such as Judge Bruce Selya of the First Circuit. But most of the names at the top are not very familiar, such as that of first-place Judge William Riley of the Eighth Circuit. Similarly, the bottom of the ranking contains some notoriously ideological judges but also contains some judges who are surprising to see at the bottom of a quality ranking. Overall, most of the judges fall into a rather large group in the middle and are not statistically distinguishable from one another.

To compare the results with those of Choi and Gulati in the context of more recent judges, consider Table III, which presents the "composite" ranking from Choi and Gulati's "tournament"[112] side-by-side with the rankings from Table II. Because Choi and Gulati ranked a much smaller number of judges, only 95 of the 383 are presented, with their ranks renumbered accordingly. Comparing the two tables suggests that there is no visible relationship between the Choi and Gulati rankings and the quality rankings as computed in this study. Some of the lowest judges in Choi and Gulati's ranking, such as Judge Stanley Marcus of the Eleventh Circuit and Judge Karen Henderson of the D.C. Circuit, come out near the top of this Article's ranking.

111. *See* Choi & Gulati, *supra* note 5; Landes et al., *supra* note 7.
112. Choi & Gulati, *supra* note 5, at 113-15 tbl.H.

Table III. Comparison with Choi and Gulati Rankings							
Columns							
1	2	3	4	5	6	7	8
Judge Name	Circuit	Score: All Panels	Rank-ing: All Panels	Ranking: Outside Panels	Ranking: All Opinions	Ranking: Outside Opinions	Ranking: Choi and Gulati
Motz	4	0.068	1	6	29	32	32
GarzaE	5	0.065	2	33	19	10	36
Wilkins	4	0.064	3	8	11	11	37
King	5	0.059	4	4	5	4	17
McKee	3	0.058	5	20	9	27	31
Hender-son	12	0.057	6	3	17	36	69
Marcus	11	0.052	7	68	41	51	74
Ebel	10	0.052	8	7	12	16	6
Ripple	7	0.051	9	36	3	15	21
WoodD	7	0.05	10	25	23	34	8
Nygaard	3	0.049	11	5	2	2	29
Lynch	1	0.048	12	15	27	28	11
Rogers	12	0.048	13	26	15	5	48
Scirica	3	0.046	14	9	36	35	25
Walker	2	0.046	15	16	13	23	26
Trott	9	0.045	16	14	55	69	18
Ca-branes	2	0.044	17	45	4	20	54
Garland	12	0.042	18	49	1	1	56
Alito	3	0.042	19	18	42	29	16
Carnes	11	0.041	20	40	35	59	7
Tacha	10	0.04	21	38	8	14	64
Gilman	6	0.04	22	1	46	19	15
Barks-dale	5	0.04	23	17	26	9	67
Lucero	10	0.04	24	12	22	6	65
Birch	11	0.039	25	10	52	45	40
Kanne	7	0.039	26	59	6	39	63
Easter-brook	7	0.038	27	39	30	31	2
Luttig	4	0.035	28	65	49	63	10
Murphy	10	0.035	29	23	21	12	33
Hull	11	0.035	30	19	14	56	50
Michael	4	0.035	31	21	57	46	66
Rymer	9	0.034	32	2	62	54	73
Boudin	1	0.034	33	52	44	25	20
Manion	7	0.034	34	43	28	30	53
Wilkin-son	4	0.033	35	41	48	49	3

Moore	6	0.033	36	29	43	50	12
Evans	7	0.033	37	13	7	7	39
Kelly	10	0.032	38	22	40	8	24
Briscoe	10	0.031	39	48	64	74	52
Williams	4	0.03	40	24	25	21	49
Hawkins	9	0.03	41	66	24	33	71
Sentelle	12	0.027	42	46	38	13	57
Rovner	7	0.027	43	42	31	18	46
Boggs	6	0.026	44	53	60	66	9
Randolph	12	0.025	45	70	20	55	47
Niemeyer	4	0.025	46	27	37	24	4
Stewart	5	0.024	47	11	18	3	35
Edwards	12	0.024	48	61	33	61	51
Edmondson	11	0.024	49	56	32	17	28
Black	11	0.021	50	50	73	62	60
Loken	8	0.017	51	51	58	38	59
Tatel	12	0.017	52	64	61	53	44
ParkerB	2	0.017	53	55	10	26	62
GinsburgD	12	0.017	54	67	53	71	41
Daughtrey	6	0.016	55	37	74	67	72
Jacobs	2	0.015	56	30	34	42	23
Schroeder	9	0.015	57	47	69	73	42
Clay	6	0.015	58	44	45	44	34
Smith	5	0.015	59	63	56	60	5
JonesE	5	0.015	60	69	47	64	13
Benavides	5	0.014	61	73	50	68	38
Barkett	11	0.014	62	54	68	72	19
Posner	7	0.013	63	58	54	57	1
Higginbotham	5	0.012	64	35	70	58	22
Henry	10	0.011	65	31	67	40	68
Dubina	11	0.011	66	32	39	37	61
Batchelder	6	0.011	67	28	71	48	70
Seymour	10	0.01	68	57	63	65	45
Cole	6	0.007	69	34	65	52	58
Rendell	3	-0.004	70	72	51	70	30
Arnold, M.S.	8	-0.004	71	60	59	47	14
Kozinski	9	-0.004	72	62	72	43	55
Thomas	9	-0.011	73	71	16	22	27

Klein-feld	9	-0.04	74	74	66	41	43

The most remarkable surprises, however, are probably the judges who ended up in the middle or bottom of this ranking but at the top of the Choi and Gulati ranking. In particular, the conspicuous absence of the influential judge *par excellence* in the Choi and Gulati ranking – Judge Richard Posner – from the top of this list is striking. In most rankings in the citation-count studies, Judge Posner ranks first among federal appellate judges, in many cases far ahead of the second place judge.[113] In Table III, however, we find that Judge Posner, although ranked first in the Choi and Gulati study, is well below the median in terms of the mix of positive and negative citations. Similarly, the judges who placed second (Frank Easterbrook), third (J. Harvie Wilkinson), fourth (Paul Niemeyer), and fifth (Jerry Smith) in Choi and Gulati's composite measure are similarly undistinguished in this ranking, all hidden in the middle of Table III. Only Judge David Ebel of the Tenth Circuit and Judge Diane Wood of the Seventh Circuit make the top ten judges on both lists in Table III. Why the striking differences between the results in Tables II and III, on the one hand, and the citation-count studies, on the other hand?

To answer this question, we must first determine why the rankings in column 4 of Tables II and III are so different from those of the citation-count studies. Is it the use of inside-circuit as well as outside-circuit decisions? Or is it the use of panels on which judges served rather than the opinions they wrote? Or both? Consider column 5 of Table III, which presents the results of the same (panel-based) model based only on outside-circuit citations. The rankings have a strong relationship to those in column 4, which is not surprising because they include many of the same data points but still bear no relationship to the Choi and Gulati rankings. It does not appear that the inside-circuit versus outside-circuit citations are driving the different results between the two approaches.

What about the use of panel membership in this study rather than only authored opinions as in citation-count studies? Column 6 conducts the same analysis as column 4 but ranks judges based on positive and negative citations to their authored opinions, rather than on the panels in which they participated. Again, the results resemble those in column 4 but differ sharply from those of Choi and Gulati. Finally, column 7 of Table III presents the opinion-based model utilizing only outside-circuit citations, making the dataset most similar to Choi and Gulati. The results still are not any closer to resembling those of the citation-count studies. Thus, the differences in rankings between

113. *See* Stephen J. Choi & G. Mitu Gulati, *Mr. Justice Posner? Unpacking the Statistics*, 61 N.Y.U. ANN. SURV. AM. L. 19, 21 (2005) ("Judge Posner far outdid his judicial colleagues by almost every measure we used."). Even in Cross and Lindquist, which is not based on citations, Judge Posner ranks rather high. *See* Cross & Lindquist, *supra* note 18, at 1410-11.

this study and the count studies result not only from the use of inside-circuit citations or panels rather than authored opinions but also from the use of positive and negative citations.

One might still argue, however, that the results in this Article do not meaningfully measure whether decisions are cited positively or negatively. For example, perhaps the Shepard's Citation treatments are not as valid and reliable as the studies have suggested and perhaps the objectivity of citation counts makes a better measure of quality. To determine whether this explanation accounts for the different results in this Article, one can examine the coefficients on variables for which we know the effect – the control variables – to learn how well the positive and negative treatments from Shepard's Citations correspond to actual judicial citations. For example, it would be very surprising if judges did not cite precedents in their own circuits more favorably than those outside their circuits, or if judges did not cite other judges of their political party more favorably than judges of the opposite party. So if the Shepard's Citations measures did not reflect these empirical regularities, we might suspect a problem with the data. But if the Shepard's Citations measures do reflect these empirical regularities, we know that they are, in fact, measuring something meaningful.

Table IV presents the results for these control variables, one column for each column of Table II. As expected, the coefficients are positive and large for inside-circuit citations (including the Eleventh Circuit citing the Fifth Circuit), indicating that inside-circuit citations are treated more positively than outside-circuit citations. Similarly, the ideological variables are also significant and in the directions that one would expect:[114] judges appointed by Republicans cite judges appointed by Democrats more negatively than they cite those appointed by Republicans, and the opposite is true as well. Most of the ideological variables are also highly statistically significant, although nowhere near the circuit variables either in magnitude or in statistical significance. These control variables, all in the expected directions and virtually all highly statistically significant, suggest that the positive and negative Shepard's treatment codes, on average, measure exactly what they purport to measure.

Table IV. Control Variables

Variable	All Panels	Outside Panels	All Opinions	Outside Opinions	By Democrats	By Republicans
Eleventh Circuit citing Fifth Circuit	0.093 (.008)	NA	0.093 (.008)	NA	0.119 (.014)	0.078 (.014)
First Circuit	0.105 (.007)	NA.	0.106 (.007)	NA	0.101 (.014)	0.11 (.01)

114. The interpretation of these coefficients is discussed in more detail, *infra* Part V.

Table IV. Control Variables (cont)

Variable	All Panels	Outside Panels	All Opinions	Outside Opinions	By Democrats	By Republicans
Second Circuit	0.044 (.005)	NA	0.047 (.005)	NA	0.031 (.01)	0.05 (.008)
Third Circuit	0.068 (.007)	NA	0.066 (.007)	NA	0.066 (.014)	0.074 (.009)
Fourth Circuit	0.125 (.008)	NA	0.131 (.008)	NA	0.144 (.015)	0.12 (.011)
Fifth Circuit	0.125 (.005)	NA	0.127 (.004)	NA	0.145 (.009)	0.114 (.007)
Sixth Circuit	0.158 (.006)	NA	0.152 (.006)	NA	0.169 (.011)	0.146 (.009)
Seventh Circuit	0.107 (.005)	NA	0.105 (.005)	NA	0.084 (.012)	0.111 (.007)
Eighth Circuit	0.138 (.006)	NA	0.146 (.006)	NA	0.131 (.012)	0.133 (.008)
Ninth Circuit	0.073 (.005)	NA	0.057 (.004)	NA	0.069 (.008)	0.073 (.007)
Tenth Circuit	0.145 (.007)	NA	0.146 (.007)	NA	0.156 (.012)	0.133 (.01)
Eleventh Circuit	0.136 (.007)	NA	0.144 (.006)	NA	0.139 (.014)	0.131 (.01)
D.C. Circuit	0.052 (.008)	NA	0.049 (.007)	NA	0.067 (.013)	0.043 (.011)
Repub citing Repub	0.007 (.003)	0.008 (0.004)	0.016 (.002)	0.017 (.004)	NA	0.004 (.005)
Repub citing Dem	-0.014 (.003)	-0.015 (0.005)	-0.007 (.003)	-0.003 (.004)	NA	-0.006 (.005)
Dem citing Repub	-0.012 (.003)	-0.011 (0.005)	-0.004 (.003)	-0.003 (.005)	-0.013 (.006)	NA
Dem citing Dem	0.012 (.004)	0.005 (0.006)	0.018 (.003)	0.017 (.005)	-0.009 (.006)	NA

**Excludes en banc panels. Standard errors in parentheses.*

B. Sitting Judges and the Sotomayor Nomination

The analysis performed above attempts (perhaps overly ambitiously) to place federal appellate judges who served at different times over a fifty-year period on a common scale of quality. This requires introducing control variables for different time periods such that, perhaps counterintuitively, some judges who are higher in the ranking can have a lower proportion of positive votes than some judges who are lower in the ranking, depending on when the judges served. To the extent that these controls are incorrectly constructed, they may distort the analysis and present biased results. Nevertheless, some control variables are necessary to compare judges intertemporally in a dataset with such strong patterns over time. Because some might object to the choice of control variables or find the methodology less than intuitive, this section

develops an alternative ranking for current judges based on simple percentages.

By comparing positive and negative citations to judges who served over the same time period, one arguably can get a rough sense of relative judicial quality without using any control variables.[115] This section uses the opportunity presented by the nomination of then-Judge Sonia Sotomayor to the Supreme Court to present such a ranking for the roughly ten-year period she served on the Second Circuit. The following two tables therefore present a ranking of the ninety-five active[116] judges who served during the same period that then-Judge Sotomayor was a circuit judge, beginning on October 7, 1998 (the date of her commission)[117] and ending in mid-2008 (the start of data collection). The judges are ranked by the percentage of positive citations (among all positive and negative citations) to the published opinions each judge wrote during the period.[118] Separate tables are presented for outside-circuit citations, a panel in another circuit citing a judge's opinion, and inside-circuit citations, a panel in the same circuit citing the judge's opinion.

Table V presents the simplified ranking based on outside-circuit citations and Table VI presents the simplified ranking based on inside-circuit citations. The results are rather surprising. Then-Judge Sotomayor ranks third out of ninety-five active judges in outside-circuit citations, with 81.6% of her citations positive. However, she ranks eightieth out of ninety-five in inside-circuit citations, with 76% of her citations positive. Indeed, then-Judge Sotomayor's inside-circuit citations are more negative than her outside-circuit citations, which is true of only five other judges in Table V. Moreover, the gap between her inside-circuit and outside-circuit ratings is larger than any of these other five judges. Even more surprising, the next largest

115. I say "arguably" because even if judges serve during the same periods, there may be factors that affect positive and negative citations other than temporal factors. For example, such an approach does not control for any differences between circuits or other factors that could dramatically affect the rankings.

116. This analysis excludes judges who have taken senior status, following the approach taken by Choi and Gulati, *supra* note 5. *See also* Choi & Gulati, *supra* note 108, at 97. Senior judges are excluded for three reasons. First, this makes the analysis consistent with Choi and Gulati for comparison. Second, senior judges are less likely to be nominated to the Court, so it makes less sense to compare them to then-Judge Sotomayor in the context of her nomination. And third, senior judges have a potentially different mix of cases and therefore may not be comparable to active judges.

117. Judge Robert King of the Fourth Circuit was included even though he received his commission two days after then-Judge Sotomayor and therefore technically did not serve the entire period. *See* Judges of the Fourth Circuit, Since 1801, http://www.ca4.uscourts.gov/pdf/HistoryJudges.pdf (last visited Feb. 9, 2011).

118. Citations to judges' opinions are used instead of panel membership because some panel members will not have served during the whole period and because the ratios are more intuitive than the results of a regression analysis holding panel membership constant.

difference of this kind is that of Judge Merrick Garland of the D.C. Circuit, who was widely reported to be one of the three top courts of appeals candidates for the nomination that ultimately went to then-Judge Sotomayor.[119] What could explain this remarkable pattern?

Table V. Outside Circuit Ranking

Judge Name	Circuit	Positive Outside	Negative Outside	Number of Opinions Cited Outside the Circuit	Outside Rating
Garland	12	23	3	21	0.885
King	4	56	12	47	0.824
Sotomayor	2	111	25	66	0.816
Stewart	5	48	11	37	0.814
Dennis	5	19	5	21	0.792
Kelly	10	53	15	49	0.779
Lucero	10	59	17	53	0.776
King	5	65	19	46	0.774
Gilman	6	61	18	62	0.772
Silverman	9	20	6	17	0.769
Wollman	8	78	24	58	0.765
Barksdale	5	32	10	25	0.762
Sloviter	3	98	31	64	0.760
Hull	11	62	22	53	0.738
Loken	8	109	39	80	0.736
Murphy	10	53	19	43	0.736
Torruella	1	77	28	70	0.733
Flaum	7	90	33	82	0.732
Luttig	4	51	19	39	0.729
McKeown	9	48	18	39	0.727
Gould	9	48	18	48	0.727
Lipez	1	77	29	65	0.726
Graber	9	61	23	43	0.726
Black	11	61	23	39	0.726
Motz	4	83	32	62	0.722
Easterbrook	7	256	99	159	0.721
Lynch	1	181	70	119	0.721
Jacobs	2	59	23	44	0.720
Dubina	11	41	16	33	0.719
GinsburgD	12	15	6	18	0.714
Hawkins	9	47	19	34	0.712
Wilkins	4	71	29	45	0.710
Cabranes	2	88	36	65	0.710

119. A number of reports suggested that the top contenders for the nomination among sitting courts of appeals judges were Judges Diane Wood of the Seventh Circuit, then-Judge Sotomayor of the Second Circuit, and Judge Garland of the D.C. Circuit. *See, e.g.*, *All Things Considered: Women Dominate Supreme Court Short List* (NPR radio broadcast May 13, 2009), *available at* http://www.npr.org/templates/story/story.php?storyId=104103838.

Kanne	7	127	53	94	0.706
Evans	7	45	19	41	0.703
Batchelder	6	45	19	30	0.703
Michael	4	59	25	42	0.702
Niemeyer	4	87	37	67	0.702
Tacha	10	75	32	56	0.701
Ripple	7	103	44	99	0.701
Traxler	4	49	21	40	0.700
Boudin	1	79	35	69	0.693
Wiener	5	46	21	41	0.687
Clay	6	83	38	66	0.686
Barry	3	26	12	17	0.684
Davis	5	41	19	34	0.683
Schroeder	9	28	13	24	0.683
Rogers	12	17	8	14	0.680
Edmondson	11	21	10	16	0.677
Henderson	12	23	11	16	0.676
WoodD	7	94	45	79	0.676
Marcus	11	82	40	43	0.672
Rendell	3	53	26	43	0.671
Anderson	11	34	17	27	0.667
Rovner	7	56	28	60	0.667
Scirica	3	72	36	47	0.667
Moore	6	115	58	101	0.665
Murphy	8	61	31	59	0.663
Wilkinson	4	110	56	85	0.663
Calabresi	2	47	24	45	0.662
Pregerson	9	45	23	41	0.662
Kozinski	9	33	17	23	0.660
Williams	4	61	32	50	0.656
Carnes	11	91	48	65	0.655
Tjoflat	11	56	30	52	0.651
Jolly	5	66	36	51	0.647
Pooler	2	42	23	48	0.646
Sentelle	12	20	11	21	0.645
Birch	11	45	25	42	0.643
Martin	6	27	15	25	0.643
McKee	3	27	15	28	0.643
OScannlain	9	50	28	47	0.641
Barkett	11	48	27	44	0.640
Cole	6	39	22	34	0.639
GarzaE	5	46	26	39	0.639
Kleinfeld	9	21	12	22	0.636
FletcherW	9	50	29	41	0.633
Thomas	9	31	18	30	0.633
Posner	7	222	131	185	0.629
Benavides	5	49	29	43	0.628
Tatel	12	31	19	31	0.620
Straub	2	26	16	28	0.619
Rymer	9	26	16	35	0.619
Henry	10	36	23	33	0.610

Table V. Outside Circuit Ranking (cont)

Judge Name	Circuit	Positive Outside	Negative Outside	Number of Opinions Cited Outside the Circuit	Outside Rating
Randolph	12	26	17	27	0.605
Smith	5	73	50	71	0.593
Sack	2	59	41	54	0.590
Daughtrey	6	31	22	29	0.585
Boggs	6	49	37	50	0.570
Fisher	9	37	28	36	0.569
JonesE	5	54	44	53	0.551
Paez	9	22	18	31	0.550
Briscoe	10	38	37	41	0.507
Wardlaw	9	19	21	26	0.475
Reinhardt	9	58	73	62	0.443

Table VI. Inside Circuit Ranking

Judge Name	Circuit	Positive Inside	Negative Inside	Opinions Inside	Inside Rating
Davis	5	149	13	59	0.920
Anderson	11	78	7	42	0.918
Murphy	8	173	19	113	0.901
Hull	11	282	32	114	0.898
Torruella	1	297	34	184	0.897
Wollman	8	260	30	163	0.897
Henderson	12	47	6	31	0.887
Marcus	11	186	26	67	0.877
Lucero	10	160	24	90	0.870
Moore	6	626	94	224	0.869
Tacha	10	205	31	89	0.869
Ripple	7	350	53	197	0.868
Kanne	7	347	55	204	0.863
Tjoflat	11	176	28	107	0.863
WoodD	7	280	46	180	0.859
King	4	91	15	68	0.858
Edmondson	11	82	14	31	0.854
Barksdale	5	99	17	64	0.853
Loken	8	273	47	139	0.853
Carnes	11	232	41	101	0.850
Clay	6	316	56	138	0.849
Birch	11	124	22	76	0.849
Rovner	7	234	42	159	0.848
Lynch	1	423	76	214	0.848
Murphy	10	154	28	92	0.846
Traxler	4	60	11	43	0.845
Gilman	6	321	59	190	0.845
Briscoe	10	103	19	68	0.844
Motz	4	59	11	47	0.843
Flaum	7	268	50	179	0.843
King	5	203	38	88	0.842
Wilkins	4	80	15	47	0.842

Cabranes	2	165	31	75	0.842
Randolph	12	69	13	37	0.841
GarzaE	5	137	27	90	0.835
Wilkinson	4	86	17	67	0.835
Wiener	5	111	22	79	0.835
Garland	12	70	14	46	0.833
Scirica	3	117	24	53	0.830
Lipez	1	219	45	140	0.830
Boggs	6	198	41	106	0.828
Boudin	1	186	39	109	0.827
Williams	5	57	12	44	0.826
Michael	4	38	8	31	0.826
Dennis	4	57	12	40	0.826
Barkett	11	75	16	56	0.824
FletcherW	9	142	31	82	0.821
Evans	7	128	28	100	0.821
Dubina	11	63	14	45	0.818
Martin	6	130	29	68	0.818
JonesE	5	118	27	76	0.814
McKeown	9	120	28	87	0.811
Easterbrook	7	268	63	191	0.810
Hawkins	9	80	19	43	0.808
Paez	9	116	28	48	0.806
Benavides	5	165	40	95	0.805
Posner	7	461	112	265	0.805
Straub	2	65	16	48	0.802
Cole	6	162	40	85	0.802
Luttig	4	52	13	40	0.800
Jolly	5	151	38	86	0.799
Henry	10	130	34	79	0.793
Kelly	10	166	44	102	0.790
Rendell	3	79	21	55	0.790
GinsburgD	12	63	17	40	0.788
Pregerson	9	137	37	75	0.787
Fisher	9	111	30	60	0.787
Rymer	9	85	23	44	0.787
Thomas	9	113	31	63	0.785
Calabresi	2	116	32	75	0.784
Wardlaw	9	94	26	54	0.783
Stewart	5	140	39	93	0.782
Gould	9	138	39	77	0.780
Graber	9	118	34	73	0.776
Black	11	83	24	57	0.776
Silverman	9	62	18	39	0.775
Daughtrey	6	92	27	52	0.773
Sentelle	12	51	15	44	0.773
Smith	5	231	68	127	0.773
Sotomayor	2	152	48	100	0.760
Barry	3	22	7	20	0.759
Schroeder	9	83	27	46	0.755
McKee	3	58	19	43	0.753

Table VI. Inside Circuit Ranking (cont)					
Judge Name	Circuit	Positive Inside	Negative Inside	Opinions Inside	Inside Rating
Reinhardt	9	219	77	102	0.740
Tatel	12	45	16	41	0.738
Niemeyer	4	90	33	72	0.732
OScannlain	9	121	45	80	0.729
Sloviter	3	120	45	83	0.727
Batchelder	6	102	39	59	0.723
Jacobs	2	74	30	65	0.712
Pooler	2	89	40	64	0.690
Rogers	12	42	19	40	0.689
Kleinfeld	9	52	24	38	0.684
Kozinski	9	47	22	40	0.681
Sack	2	71	34	70	0.676

One possible explanation for the discrepancy could be the widely held perception that both the Second and D.C. Circuits are highly influential and prestigious outside their circuits[120] (thus accounting for the judges' high rankings for outside-circuit citations) and less deferential toward their own precedents than most other circuits (thus accounting for the judges' low ranking for inside-circuit citations). The data presented earlier,[121] however, combined with the results presented in Part V.D.1 below, suggest that although the second assertion is right, the first one is not – the Second Circuit and D.C. Circuit judges have no higher outside-circuit ratings on average than judges in most other circuits. Another possible interpretation is that Judge Garland and then-Judge Sotomayor were writing for a national audience, emphasizing aspirational legal principles, rather than following precedents within their own circuits. Finally, it seems possible that collegiality or other interpersonal considerations enter into *intra*-circuit citations to a larger extent than they enter into *inter*-circuit citations and could explain some of the discrepancy between the two measures.[122]

Whatever the interpretation of these tables, it appears that then-Judge Sotomayor's work product on the Second Circuit is highly respected in the federal judiciary, particularly outside the Second Circuit. So how does she compare to historic and current judges in the original model with control variables and the larger dataset? Returning to Table II, we can see that the model suggests then-Judge Sotomayor is among the most positively cited judges since 1960. In every subset of the data, then-Judge Sotomayor is in the top 31 of the 383 ranked judges, and actually ranks first among all such

120. Solimine, *supra* note 61, at 1341-42 ("Over much of the twentieth century, the conventional wisdom has ranked the Second and District of Columbia Circuits highest with regard to prestige and influence.").

121. *See supra* Part IV.A.

122. *See infra* Part IV.C.

judges in outside-circuit citations to panels on which she served.[123] Indeed, as will be discussed in more detail below, columns 8 and 9 of Table II show that then-Judge Sotomayor received virtually the same rating from judges appointed by Democrats and those appointed by Republicans. These results reveal a great deal about her ability to produce persuasive consensus opinions that command respect throughout the federal appellate judiciary.

The illustration of then-Judge Sotomayor's nomination to the Supreme Court is an example of how the analysis in this Article makes a practical contribution that complements other judge ranking techniques. This study ranks then-Judge Sotomayor among the most outstanding judges since 1960, coming in at seventeenth in the primary measure of Table II and first according to another measure. Although then-Judge Sotomayor was not ranked in the Choi and Gulati study,[124] there is reason to believe that she would rank high – although not as high – according to those measures as well.[125] Thus, then-Judge Sotomayor, together with Judge Bruce Selya and a handful of others, is among those relatively unusual judges who perform well according to both measures.

C. Interpretation

As discussed above, the rankings of judicial quality presented in this Article differ substantially from the Choi and Gulati studies and from the rest of the citation-count literature. Many of the highly ranked judges of the citation-count studies are relegated to mediocrity in this approach. Similarly, most of the top judges in this study are not the well-known names found in the Choi and Gulati studies. How should one distinguish these scores from those presented in Choi and Gulati? Is one of the approaches to measuring judicial quality more accurate than the other, or are the scores better viewed as measuring different dimensions of judicial performance?

The approach in this Article can be thought of as emphasizing dimensions that are not directly measured by citation-count studies – such as traditional legal reasoning, collegiality, and careful analysis – and de-emphasizing dimensions highlighted by the citation-count studies – such as creativity, originality, and entrepreneurship. In this sense, the two measures really are complements, as both capture potentially important qualities in a judge and

123. The reason Judge Merrick Garland rates lower on this measure is that the Table II measure is based on panels, not opinions, as was the case in Table V. Judge Garland ranks higher in Table II according to citations to his opinions. Judge Robert King was not included in Table II because the standard error for his coefficient was too large. He would have ranked second according to that measure, however, outranking both Judge Garland and then-Judge Sotomayor.

124. *See* Choi & Gulati, *supra* note 5, at 113-15 tbl.H.

125. Eric Posner has conducted a rough count for a blog posting. *See* Posting of Eric Posner to the Volokh Conspiracy, http://volokh.com/posts/1243482653.shtml (May 27, 2009, 23:50 EST).

use a similar methodology, even though they produce radically different assessments of the performance of federal appellate judges.

The idea that a quantitative methodology could measure intangible qualities like collegiality is certainly controversial. But what is probably not controversial is that to the extent collegiality is measured by citation analysis, this study measures collegiality better than the citation-count studies. The first reason is that inside-circuit citations are more likely to assess collegiality, temperament, and the like than are outside-circuit citations, and this study places equal emphasis on both inside- and outside-circuit citations. Second, the use of panel membership, rather than opinion authorship, to link judges to citations may allow this analysis to capture collegiality in a way that citations to authored opinions do not. This is because linking judges to cases through panel participation rewards judges who positively contribute to a joint, rather than an individual, intellectual product. This idea of contributing to a joint intellectual product is much of what collegiality in appellate courts is (or should be) about.

This potential to capture intangible aspects of collegiality in a quantitative measure may be useful for Supreme Court appointments, as commentators have identified the ability to influence other members in a group decision-making situation as an important part of success at the Supreme Court level.[126] The argument is that "[j]udges in conference who carefully engage and evaluate the arguments and explanations offered by their colleagues are likely to create a better informed and intellectually more rigorous final product even if no one's vote is changed during deliberations."[127] This "final product" is exactly what other judges evaluate when they decide to cite a decision positively or negatively. Therefore, to the extent a judge contributes positively to that "final product" when he or she is not writing the opinion, the judge exhibits the valuable traits that are often referred to as collegiality. As a result, the use of panel membership in this study arguably reflects some of the intangible qualities of good judges that are often collectively referred to as "collegiality."[128] Thus, while quantitative approaches may never supplant more traditional means of evaluating intangible qualities, such as the ABA evaluation for integrity and judicial temperament,[129] it may be that quantitative approaches can capture some intangible qualities as well.

---

126. Brudney, *supra* note 71, at 1045-47.

127. *Id.* at 1045.

128. *See id.* at 1046-47.

129. The ABA evaluates candidates in terms of integrity, professional competence, and judicial temperament. American Bar Association, *American Bar Association Standing Committee on Federal Judiciary: What It Is and How It Works*, 17 J.L. & POL. 215, 215 (2001). Arguably, the comprehensive approach to judicial evaluation in this Article evaluates the judicial competence of a federal appellate judge much more fairly and systematically than does the ABA process. The approach in this Article has little or nothing to say about a judge's integrity or temperament, ex-

The second dimension that is captured better by this approach than the citation-count studies is a judge's ability to persuasively argue from traditional legal materials and engage in traditional legal reasoning. This dimension is, in a sense, an alternative to the qualities the citation-count studies emphasize – creativity, originality, and entrepreneurship – that are in some ways the opposite of traditional legal reasoning. As Cross and Lindquist observed, "[t]here is reason to think that the Choi and Gulati results capture an inclination toward judicial entrepreneurship. One might expect such judges to be more frequently cited because they strive to make the law."[130] Similarly, another commentator notes that the citation-count approach "might encourage judges to use novel reasoning or reach atypical conclusions so as to increase the number of citations to their opinions."[131]

The willingness or reluctance of individual judges to aggressively "make the law" appears to be an important part of what the rankings in this Article are measuring. As an illustration, cases in which the appellate panel reverses a lower court decision are cited much more negatively than cases in which the appellate panel affirms a lower court decision.[132] Indeed, this effect is considerably larger than the effect of the political parties of the presidents who appointed the judges, or almost any other single variable in the analysis. To the extent that a reversal of a district court tends to be more aggressive or entrepreneurial than an affirmance, this pattern supports the interpretation above.

The criticism that emerges from the emphasis on creativity and novelty is that "novel opinions are not necessarily better opinions."[133] Gulati, Choi and (Eric) Posner summarize this criticism (with which they disagree) as follows:

> Citation counts, the argument goes, likely reward judges who are more creative and expansive in their articulations of the law, since courts are more likely to cite such articulations. If anything, high

---

cept to the extent those qualities are known to and taken into account by federal judges in citing one another.

130. Cross & Lindquist, *supra* note 18, at 1421.

131. Ahmed E. Taha, *Information and the Selection of Judges: A Comment on "A Tournament of Judges"*, 32 FLA. ST. U. L. REV. 1401, 1408 (2005).

132. When the analysis in this Article was augmented with data on whether the cited decision reversed or affirmed the lower court, the results suggested that reversals were cited much more negatively than affirmances. This is consistent with what Frank Cross has found in his work on the courts of appeals. See FRANK B. CROSS, DECISION MAKING ON THE U.S. COURTS OF APPEALS 214-15 (2007). I would like to thank Frank Cross for the suggestion to test the effect of the disposition of the lower court on the citation of the appellate opinion.

133. Taha, *supra* note 131, at 1408.

> citation counts may be a measure of bad judging if the "better" approach to lawmaking is to decide cases narrowly.[134]

This is a very fair description of the criticism, but their response is ultimately unpersuasive. They respond with the economic argument that "if judges value minimalism, then minimalist opinions will be cited more, not the creative and expansive ones."[135] As explained above, this may make sense if the judges are cited positively, but expansive opinions are particularly likely to require distinguishing so that, while they receive many citations, they are negative citations.[136] Again, judges do not necessarily cite opinions because they "value" them; they often cite them because the litigants' briefs rely on them.

In contrast to the citation-count approach, which may "encourage judges to use novel reasoning or reach atypical conclusions,"[137] this Article evaluates judges' ability to use traditional reasoning to reach predictable conclusions. The Choi and Gulati rankings may favor more aggressive judicial theorizing – judges who have sweeping, encompassing theories of how cases should be resolved but not necessarily the time or inclination to work out all of the details. The Choi and Gulati approach also rewards judges only for the opinions they actually author, implicitly depreciating the panel's other judges' contributions in producing a well-reasoned decision. The approach in this Article takes a different perspective, rewarding judges for a combination of precise analysis, careful decision-making, and collegiality, whether the judge wrote the opinion or simply participated in the joint intellectual product of an appellate panel.

The comments in this section are not intended to diminish the value of the citation-count approach but rather to emphasize the complementarity of the two approaches. Part of the problem with the citation studies is that, although they use similar methods, they variously refer to their measurements as "prestige," "influence," "reputation," or "quality," typically without defining these terms or elucidating the distinctions between them.[138] As Cross and

---

134. Choi et al., *supra* note 37, at 1322 (footnote omitted).

135. *Id.*

136. The measures in this Article may actually favor judicial maximalists – those who write highly general opinions. This is because "distinguishing" citations are negative, and it may be more difficult to distinguish a highly general opinion. Frederick Schauer, *Precedent*, 39 STAN. L. REV. 571, 594 (1987) ("Although it will always be *possible* to distinguish a precedent, this becomes comparatively harder if we describe and use the precedents of the past in general terms."). But this effect is more likely to appear within the circuit than outside the circuit. This is because outside-circuit citations can simply reject ("criticize" in the language of Shepard's) the broadly-worded opinion. This must be left for further research.

137. *See* Taha, *supra* note 131, at 1408.

138. A notable exception is Klein and Morrisroe, who define "influence" as "the extent to which the actions of one person have an effect on the views or behavior of

Lindquist point out, "the Choi and Gulati quality measures may implicitly presume that judicial entrepreneurship is associated with judicial quality, a presumption they do not support."[139] The characteristics measured by the Choi and Gulati approach, including influence, prestige, productivity, creativity, and the like are valuable traits in federal appellate judges and worthwhile to measure, but they are not the same as judicial quality. Although it is unclear whether a citation study can distinguish between influence and prestige,[140] a citation study can distinguish to some extent between influence and quality. Both of these traits may be relevant to evaluating judges in certain contexts, and while the count studies focus on influence, this study focuses on quality.

V. CRITICISMS OF QUANTITATIVE ASSESSMENT AND NORMATIVE APPLICATION

The use of quantitative methods to evaluate judges is highly controversial, as the commentators' responses to the Choi and Gulati study demonstrate. Although some of the commentators raised specific methodological objections to Choi and Gulati's work, many commentators' responses were more of a general demurrer, dismissing the very idea of quantifying judicial performance.[141] Indeed, there appears to be a relatively large group of scholars opposed to the very idea of quantification of judicial performance. But the increasing availability of data on the courts means that judicial quantification projects are here to stay.[142] This Part, therefore, attempts to address some of the general criticisms of the project of evaluating judges with quantitative data. It also tries to show how quantitative assessment might be used in ways other than Supreme Court nomination and confirmation, which is where the most controversy seems to have arisen.

A. Three Types of Criticisms of Citation Studies

The skeptics of the Choi and Gulati and other citation approaches primarily advance three related arguments as a criticism of ranking judges with

---

others" and "prestige" as "the amount of respect, regard, or esteem one enjoys among one's fellows." Klein & Morrisroe, *supra* note 36, at 371-72.

139. Cross & Lindquist, *supra* note 18, at 1421.

140. *See* Smyth & Bhattacharya, *supra* note 38, at 236.

141. Apparently this type of criticism is not new; it accompanied the older studies as well. *See, e.g.*, Greely, *supra* note 31, at 107 (describing the furor associated with Posner's use of citation analysis to study Benjamin Cardozo).

142. Baker et al., *supra* note 5, at 1646 ("Judicial rankings, for better or worse, are not going away. Therefore, there is a strong need to develop meaningful and appropriate measures for evaluation.").

citations.[143] The first is that even if the Choi and Gulati studies measure quality at the appellate court level, good appellate judges do not necessarily translate into good Supreme Court justices. The second is that even if the rankings measure good appellate judges and good appellate judges do make good Supreme Court justices, ranking the judges may cause judges to change their behavior to "game" the system[144] or otherwise produce undesirable consequences. The third is that the ranking system may provide "cover" for ideological appointments instead of avoiding them.[145] This Part addresses each of these criticisms in turn.

The first of these claims is that appellate quality may not translate into Supreme Court quality. Supreme Court justices and federal appellate justices do different jobs, and scholars have argued that successful appellate judges do not necessarily make "great" Supreme Court justices.[146] James Brudney illustrates this in his response to Choi and Gulati's article by using an analysis similar to that of Choi and Gulati to compare Justices Burger and Blackmun as appellate judges. He concludes that then-Judge Burger looked better than then-Judge Blackmun on most of Choi and Gulati's measures,[147] whereas most scholars see Blackmun as the greater justice.[148] Similarly, analysis in this Article sees Blackmun (ranked 59 in column 4) and Burger (ranked 66) as essentially the same in almost all areas, with the exception of the ranking by opinions, where Blackmun excelled (ranked 5) and Burger struggled (ranked 299). But quantitative studies would have had difficulty predicting the different trajectories of the two men's lives.

Comparing the two approaches, these objections are probably at least as applicable to the analysis in this Article as to the citation-count approach, because this analysis is more calibrated to measure a judge's quality *as an appellate judge*. The Supreme Court has a discretionary docket, which has the result that in the Supreme Court "the impact of precedent is diminished, and the importance of collegial interaction increased" relative to the appellate courts.[149] But this fact cuts both ways in evaluating the usefulness of this Article's approach for evaluating potential nominees for the Supreme Court. On the one hand, because this Article's approach probably measures adherence to precedent more than the citation-count approach, the approach here is perhaps less suited to evaluating potential Supreme Court nominees than the citation-count approach. However, because this Article probably meas-

143. Indeed, these are the three organizing themes of the response by Judge Bybee and Thomas Miles. *See* Bybee & Miles, *supra* note 74, at 1058.

144. *See, e.g.*, *id.* at 1070-72; Marshall, *supra* note 9, at 122-23.

145. Marshall, *supra* note 9, at 130-32.

146. Steven Goldberg, *Federal Judges and the Heisman Trophy*, 32 FLA. ST. U. L. REV. 1237, 1241-42 (2005).

147. Brudney, *supra* note 71, at 1028-29.

148. Duane Benton & Barrett J. Vahle, *The Burger-Blackmun Relationship: Lessons for Collegiality from the Blackmun Papers*, 70 MO. L. REV. 995 (2005).

149. Brudney, *supra* note 71, at 1037-38.

ures "collegial interaction" more effectively than the citation-count approach, the approach here might be more appropriate for evaluating potential Supreme Court nominees.

Moreover, even though the jobs of appellate judges differ from those of Supreme Court justices, the ability to do one well is probably still an indication of an ability to do the other well. The jobs are not *that* different, after all, and unless one thinks that a particular person is innately suited to one court or the other, good performance at the appellate level may just mean understanding and abiding by the different roles that judges have at different levels. When the judge transitions to the role of a Supreme Court justice, he or she transitions to a new role that measures performance criteria differently. It may be that no quantitative method could predict that Justices Burger and Blackmun would part company as they did but most likely neither could any qualitative method, so the criticism really does not have much persuasive force.

The second criticism – that rankings will encourage attempts to game the system or otherwise create perverse incentives – is addressed separately by most of the commentators but is analytically very similar to the first criticism. Under this view, judges will change their behavior to focus on the superficial measures used in the rankings, shifting effort away from unobservable but high-value activities toward observable but low-value activities. Specifically, judges could potentially increase their rankings in the Choi and Gulati framework by publishing more of their decisions, whether or not those decisions add anything to the law, and by writing more concurring and dissenting opinions.[150] Although this is sometimes considered "gaming the system" or creating bad incentives,[151] the two are really the same. The very idea of "gaming the system" means having incentives behave in ways that are positive in the rankings but otherwise undesirable. Conversely, if the change in behavior were desirable, it would not be called "gaming the system," but rather "responding to positive incentives to perform well," which is part of what Choi and Gulati hoped to encourage.

Whatever the incentive problem is called, the criticism is basically the same. Judges would be stimulated to behave in undesirable ways, particularly along the "productivity" dimension of the rankings, because that is "the easiest dimension to manipulate."[152] This could produce a situation where, as Professor Solum put it: "A tournament of judges would be won by judges who possess arbitrary luck and the vices of originality and mindless productivity; the contest would be lost by those who possess the virtues of justice and wisdom. The judicial selection process should not be transformed into a

150. *See, e.g.*, Marshall, *supra* note 9, at 122-23.

151. *See* Baker et al., *supra* note 5, at 1662-64; Bruce M. Selya, *Pulling From the Ranks?: Remarks on the Proposed Use of an Objective Judicial Ranking System to Guide the Supreme Court Appointment Process*, 32 FLA. ST. U. L. REV. 1281, 1292-95 (2005); Solum, *supra* note 43.

152. Bybee & Miles, *supra* note 74, at 1070.

game."[153] Although Solum's criticism about gaming the system may have some leverage as a criticism of Choi and Gulati's approach, it really would not apply to the approach in this study. Productivity counts for nothing in the quality ranking developed here, eliminating one of the major criticisms of the Choi and Gulati approach. In addition, as already discussed, the "vice of originality" is most likely a handicap, not an advantage, when judges are ranked on the basis of the mix of positive and negative citations.

The productivity ranking is not the only manipulable component of the Choi and Gulati formula, however. The "quality" ranking is also manipulable as Judge Selya points out:

> Any judge worth his salt will tell you that there are ways to write opinions that make citation more likely. Judges can boost citation rates by writing longer opinions, publishing opinions that would otherwise go unpublished, eschewing quotations, taking controversial positions, or reaching for novel issues at the margin of a case. Accordingly, frequency of citation sometimes may signal better strategy rather than better quality.[154]

Professor Solum identifies a similar concern, arguing that under Choi and Gulati's approach, "[j]udges will have an incentive to change the law, because an opinion that makes new law – especially new law on a topic that arises frequently – is much more likely to be cited than an opinion that merely restates existing law."[155]

Although these are certainly risks in the citation-count framework, they do not apply to the ranking in this Article. Each of the proposed manipulations would actually work *against* the judge's ranking in this study.[156] A judge who "tak[es] controversial positions," "reach[es] for novel issues," or "makes new law"[157] will in fact generate many citations – including many positive citations – but will tip the *balance* toward negative, rather than positive citations. That balance is exactly what is measured in this Article. Moreover, the approach taken here, as opposed to that in Choi and Gulati, actually elevates some of Professor Solum's "judicial virtues" as the prime measurements of quality for judges.[158] The rankings developed here are particularly likely to reflect what Solum calls "diligence," "carefulness," and "judicial craft," favoring the judge who "follows the law rather than makes

153. Solum, *supra* note 43, at 1366.

154. Selya, *supra* note 151, at 1290.

155. Solum, *supra* note 43, at 1396.

156. The "longer opinion" factor is a possible exception, as it is not clear which way this factor works. It would seem longer opinions have a higher chance of being cited for something, but it is unknown whether they would be cited more positively or more negatively.

157. Choi & Gulati, *supra* note 5, at 113-15 tbl.H.

158. *See* Solum, *supra* note 43, at 1376-85.

it."[159] As a result, the incentives created by the rankings in this study – to be more positively cited than negatively cited – are not only more difficult to manipulate than citation counts but also create more positive incentives, even as interpreted by critics of Choi and Gulati.

The third criticism commonly leveled against Choi and Gulati's rankings is that the rankings will actually provide "cover" for such appointments rather than eliminate ideological appointments.[160] The risk is that the appearance of objectivity provided by a ranking system might allow a president to appoint a judge who is highly ideological, stifling the criticism that would otherwise attend such a nominee. It turns out, however, that, at least when using the ranking developed in this Article, ideological judges do not score very high. Indeed, there appears to be a trade-off between ideology and quality as measured here that makes this criticism inapplicable to the rankings in this Article. This trade-off is developed more fully in Part V.D.1 below.

Finally, in assessing these criticisms, it is important to keep in mind the relevant standards to which one would compare a judicial evaluation approach: first, what is the alternative to the use of quantitative assessment in Supreme Court appointments, and second, what is the ideal for such appointments? The alternative is essentially the status quo. Because the public has no realistic way to evaluate judicial candidates, positions on "hot-button" issues have been the test:

> In the end, the real mettle of a potential nominee to the Court lies in her opinions and character. Reading opinions (much less discerning character), however, is time-consuming and, hence, costly. It is, moreover, an inexact science. For members of the public without legal training, comprehending the often complex legal analysis of a judicial opinion is prohibitive. The public has turned instead to a less costly means of evaluating judicial nominees – looking at a nominee's positions on what Professors Choi and Gulati call the "hot button issues," such as abortion, gun rights, affirmative action, and capital punishment, among others.[161]

These are important issues, and the argument is not that the candidate's position on these issues is unimportant or irrelevant. The point is that these issues really serve as a proxy for something that is harder to measure – the prospective nominee's opinions and character, which are too difficult to directly observe.

Because the alternative to quantitative assessment is largely the status quo, this requires detailed inquiry into a small number of high-profile opi-

159. *Id.* at 1374-76, 1393.

160. Selya, *supra* note 151, at 1286 (arguing that the "ranking system may serve to provide an impenetrable cover for an essentially ideological choice").

161. Bybee & Miles, *supra* note 74, at 1056.

nions. But what is the ideal? The above quotation suggests that ideally, a group of intelligent, objective experts would read each of the judge's opinions, assess his or her other contributions to panel decisions, evaluate these opinions and contributions honestly and thoroughly, and produce a concise report that gives a comprehensive view of the nominee's body of work as a whole. The problem is that such an assessment would be too expensive, even in the context of a specific nominee for a Supreme Court vacancy. But this Article argues that such laborious, costly work already has been done by the experts who know the candidates and their work best – the judge's colleagues on the bench. Other appellate judges evaluate precedent every day and encode their evaluation into their citations to other judges' work product. In turn, these evaluations themselves are cited and therefore evaluated in later opinions, with the evaluation encoded in a positive, negative, or neutral citation. The only remaining task is to decode the data.

B. The More General Objection to Judicial Rankings

The three types of objections to citation analysis described above each have some relevance to the citation-count approach, and each is mitigated to some degree by the additional information supplied by the approach in this study.[162] But one thing all of these objections have in common is that they relate to using the rankings for Supreme Court nomination and confirmation – the primary focus of Choi and Gulati's approach.[163] But the task of Supreme Court appointments is not the only, or even the primary, reason one might want to evaluate appellate judges. This section attempts to explain why appellate judges are especially good candidates for quantitative assessment.

The idea of rankings, especially rankings of judges, seems to make legal scholars uneasy. There are, to be sure, serious problems of measurement in assessing judicial performance, and one has to be mindful of the potential perverse consequences of such quantification, which have been amply catalogued in the references cited above. But as Choi, Gulati, and Posner point out:

> [T]he same measurement difficulty exists in many employment settings – doctors, nurses, lawyers, bankers, architects, policemen, baseball umpires, etc. – and employees in these professions are regularly evaluated by their employers. If employers can measure the performance of those professionals, why should the public not be able to measure the performance of judges?[164]

---

162. *See supra* Part V.A.

163. *See* Choi & Gulati, *supra* note 5.

164. Stephen J. Choi et al., *Are Judges Overpaid? A Skeptical Response to the Judicial Salary Debate*, 1 J. LEGAL ANALYSIS 47, 64 (2009).

There is nothing particularly special about evaluating judges, and the problems of evaluating performance and the incentives created thereby are ubiquitous. Employees are regularly evaluated in both public and private sector employment, and every employee evaluation scheme is imperfect, overlooking positive behavior, failing to detect undesirable behavior, and creating incentives to emphasize the observable aspects of performance even when the unobservable aspects are more important. In this respect, there is nothing special about judges.

There are *some* special characteristics about federal appellate judges, however, although most of them actually argue in favor of applying quantitative empirical approaches to assess their work. The work performed by appellate judges is uniquely suited to quantitative study, because federal appellate judges are a large group of high-level government employees who all do approximately the same job. Of course, the Senate and House of Representatives largely fit that description and have themselves been the subject of extensive quantitative analysis,[165] but there is an additional reason to use quantitative data to assess the performance of federal appellate judges. Most judges, unlike members of Congress, purport to be doing something objective – following the law. When members of Congress are out-voted, people say the majority disagreed with them. When a trial court is reversed, however, people say it was wrong. If appellate judges purport to perform an objective function when reviewing lower courts, then they should be evaluated in objective terms.

C. Some Criticisms Not Addressed by This Study

There are many potential criticisms of citation studies that this study does not address. Two in particular, however, arise frequently and share common underlying features, so they are addressed here. The first common criticism of not only citation studies but also of all empirical studies of judicial decision-making is the failure to distinguish between "important" cases and "ordinary" cases.[166] This study treats the garden-variety ERISA case the same as the titillating substantive due process case. The cases are weighted in the data according to how often they arise in judicial citations, not according to how much debate they provoke in law review articles. This means that the weight of each case in the analysis reflects the extent to which the opinions are useful to federal judges in deciding cases, not the extent to which they are interesting to law professors.

Although some commentators might see this emphasis as a weakness, it was, in fact, a deliberate choice in the design of this study. The purpose of this Article is to assess the quality of the average or typical published opinion

165. *E.g.*, KEITH T. POOLE & HOWARD ROSENTHAL, CONGRESS: A POLITICAL ECONOMIC HISTORY OF ROLL-CALL VOTING (1997).

166. Bybee & Miles, *supra* note 74, at 1060.

rendered by a judge, not the high-profile case that is already adequately scrutinized by legal scholars and the press. Of course, this study could easily be adapted to emphasize controversial cases by subsetting the data into cases with a dissenting opinion, for example. Such cases would tap more directly into the ideological nature of decision-making, as the scope of the data would be limited to cases in which there is already disagreement. But such an approach likely would not measure judicial quality as well as that used in this study and accordingly must be left for future research on judicial ideology.

A second common criticism of citation studies is actually the opposite of the first: that the study relies too exclusively on published opinions, when most dispositions in the federal courts of appeals are unpublished.[167] This study, like the existing citation-count studies, uses only published opinions. But one might quite justifiably argue that published opinions are a biased sample of all federal court decisions and therefore are not representative of the whole corpus of federal judicial output. It is possible that judges might hold back their lower quality opinions, which could bias the study. It is also possible that the published opinions reflect only the more significant opinions, which may also be the more controversial ones. In either case, the sample is not representative of the underlying population of legal cases.

This latter criticism is a valid one, and it reveals a general problem with empirical studies of the federal courts. Nevertheless, the criticism only applies when one is attempting to draw inferences about the underlying population of *case dispositions* in the federal system, not when one is attempting to draw inferences about the *precedents created* in those dispositions. Indeed, the very idea of a study of citations to unpublished opinions is, in a sense, a contradiction in terms, because unpublished decisions are generally not citable as precedent.[168] Thus, to the extent one wishes to study the quality of the legal reasoning that forms part of the body of precedent for later cases, the unpublished decisions are not relevant. However, to the extent that one wishes to draw inferences about the dispositions of individual cases by the federal courts, citation analysis may prove less useful than other techniques.

D. Examples of Applications

The measures developed in this Article have a number of applications in both politics and scholarly research. One application has already been mentioned, that the measures can serve as one piece (although an important piece) of information in nomination and confirmation to the Supreme Court. If used for this purpose, the measures in this study should be used together with other quantitative data (such as the Choi and Gulati studies) and qualitative infor-

167. *See, e.g.*, Edwards & Livermore, *supra* note 75, at 1922-27.

168. New Federal Rule of Appellate Procedure 32.1 modifies this rule. *See* FED. R. APP. P. 32.1. The extent to which this new rule will have practical changes in citation practices remains uncertain.

mation that bears more directly on integrity, temperament, collegiality, and the like.

In addition to nomination and confirmation of justices, however, the measures developed in this Article can provide tools for analyzing research and policy questions that otherwise might elude quantitative analysis. Some research questions would use the measures of judicial quality directly as the explanatory variables of interest, as would the question of whether the Supreme Court grants certiorari more often when low-quality judges are on the panel than when high-quality judges are on the panel.[169] Other research questions would use the quality ratings as control variables when attempting to isolate the effect of other causal factors, such as the effect of ideology on decision-making. Although most of these questions must be left for future research, this Article briefly outlines two illustrations of how these measures can inform debates about the federal judiciary. First, this section examines the implications of this study for the endless debate about whether ideology affects the decisions of courts of appeals judges. Second, it shows how this study can have implications on questions of judicial administration, such as the question of splitting the Ninth Circuit.

1. Ideology and Quality in the Courts of Appeals

The ideological variables in this model suggest that ideology (as measured by party of nominating president) does play a role in the citation of cases; judges appointed by Democrats cite those appointed by Republicans more negatively than they cite those appointed by Democrats, and the opposite is true. The effect is highly statistically significant in each category, but the effects are relatively small compared to the inside-circuit effects and even some of the individual judge coefficients. To the extent that the party of the appointing president captures ideology, these measures suggest that party plays a smaller role than inside-circuit precedent and even plays a smaller role than judicial quality in most cases. These findings are consistent with those of Frank Cross's comprehensive empirical analysis of the federal appellate courts, where he concludes that ideology has a statistically significant effect but a relatively modest role in the decisions of the courts of appeals.[170]

---

169. Landes, Lessig, and Solimine say this is so, suggesting the possibility that the Supreme Court rarely takes cases decided by our top-ranked judges because these judges not only get things "right" but their decisions are more likely to be followed in other circuits. But when the Court takes a case by a topranked [sic] judge, it is more likely to reverse his decision in order to prevent the "error" from spreading to other circuits.

Landes et al., *supra* note 7, at 325-26. This is an "open question." Cross & Lindquist, *supra* note 18, at 1404.

170. *See* CROSS, *supra* note 132, at 38 (concluding that although "[t]he results are fairly consistent in showing some effect of ideology that is typically a statistically

Of course, the party of the appointing president is a highly imperfect measure of ideology, so the results establish a floor for the degree of partisanship in appellate citations, not a ceiling.[171] Although this finding is very robust, it is essentially nothing more than a rehashing of the interminable law-political science debate about whether law or ideology affects judicial decision-making. In the federal appellate courts, the answer is both.

The results do support some novel and interesting findings about ideology. Recall that one of the key assumptions in this Article is that inside-circuit and outside-circuit citations are both useful for ascertaining the effects of judicial quality and ideology. This stems from a mild skepticism about the constraining force of precedent in guiding appellate decisions. In contrast, Choi and Gulati (echoing Landes et al. before them) emphasize outside-circuit citations on the ground that inside-circuit citations will be dictated by precedent and therefore will not be as useful for differentiating among judges.[172] If precedent constrains judges' ability to favor well-reasoned opinions within the same circuit, we would expect that precedent would also constrain the effect of partisanship and ideology; but comparing the political party coefficients of columns 2 and 4 with 3 and 5, respectively, we see that political party matters about the same amount in inside-circuit and outside-circuit citations. It would appear that to the extent precedent constrains within the circuit, it does not constrain ideology.

To take a closer look at this phenomenon, consider the results of the first (panel-based) model where the party variables are replaced by a simple variable for whether the citing case and the cited case were of the same party or different parties and how that variable interacts with the "inside-circuit" variable for each circuit. The results of this model are presented in Table VII.

Table VII. Ideology by Circuit

Variable	*Coefficient*	*P-Value*
Same Party	0.02	0.000
Eleventh Circuit citing Fifth Circuit	0.108	0.000
First Circuit	0.103	0.000
Second Circuit	0.049	0.000
Third Circuit	0.086	0.000
Fourth Circuit	0.134	0.000
Fifth Circuit	0.129	0.000
Sixth Circuit	0.159	0.000
Seventh Circuit	0.094	0.000
Eighth Circuit	0.118	0.000

significant association. . . . the measured effect size for ideology is always a fairly small one").

171. Although, the use of exclusively published opinions in this study may overestimate the degree of ideological disagreement if published opinions differ systematically from unpublished ones along this dimension, as Judge Edwards contends. *See* Edwards & Livermore, *supra* note 75, at 1950.

172. *See* Choi & Gulati, *supra* note 5; Landes et al., *supra* note 7.

Ninth Circuit	0.063	0.000
Tenth Circuit	0.14	0.000
Eleventh Circuit	0.121	0.000
D.C. Circuit	0.058	0.000
Eleventh Circuit citing Fifth Circuit * Same Party	-0.002	0.913
First Circuit * Same Party	-0.003	0.82
Second Circuit * Same Party	-0.011	0.27
Third Circuit * Same Party	-0.035	0.007
Fourth Circuit * Same Party	0.004	0.823
Fifth Circuit * Same Party	-0.006	0.451
Sixth Circuit * Same Party	-0.006	0.565
Seventh Circuit * Same Party	0.019	0.056
Eighth Circuit * Same Party	0.017	0.14
Ninth Circuit * Same Party	0.027	0.001
Tenth Circuit * Same Party	0.001	0.917
Eleventh Circuit * Same Party	0.002	0.902
D.C. Circuit * Same Party	0.004	0.747

The key information is in the first variable (Same Party) and the bottom half of the table, where the interaction effects are listed. The coefficient on Same Party shows quite clearly that citations are more positive when the citing panel and the cited panel are of the same party. That much is consistent with the existing literature. The second half of the table, however, shows that in most circuits, political party has the same effect on inside-circuit citations as outside-circuit citations. This is a surprising finding, because one might expect partisanship to be dampened in the context of inside-circuit citations where precedent (or at least the threat of en banc review) should constrain. Even more surprising is that in the Ninth Circuit, the effect of party is even stronger on inside-circuit citations than outside-circuit citations. Only in the Second and Third Circuits are inside-circuit citations significantly less partisan in nature than outside-circuit citations. In most circuits, there may be a slightly greater influence of party on inside-circuit citations, although no others rise to the level of statistical significance.

How does this finding square with the idea that precedent is binding within a circuit, and what are the implications for judicial quality? The evidence suggests a somewhat troubling state of affairs where partisanship is not significantly mitigated in most circuits, even when precedent should be a constraint. The implications of this puzzling phenomenon for the relationship of precedent to ideology must await further research, but the implications of this finding are clear for researching judicial quality: if ideological differences are able to persist in inside-circuit citations despite the constraining force of precedent, it seems even more likely that differentials in judicial quality are able to withstand that force. Indeed, even when the data are limited to inside-circuit citations, there is a considerable spread in judicial quality estimates. Thus, this evidence confirms the decision to use inside-circuit as well as outside-circuit citations to assess quality.

The measures of ideology used in this Article are rough approximations, however, which means that the analysis cannot control completely for the

effects of ideology. As a result, it is possible that the quality measures are to some extent influenced by ideological differences among the judges. Several of the responses to Choi and Gulati raised this concern, which was expressed by one comment as follows:

> If, as many argue is the case, the federal judiciary is becoming increasingly conservative, then conservative judges may have higher citation rates simply because they have more like-minded colleagues on the bench than their liberal counterparts. If this is the case, citation counts may be a systematically biased measure: judges in the ideological majority will always tend to be cited more often, and thus win this round of the tournament.[173]

This response goes on to argue that Choi and Gulati's approach may suffer from this problem: "As it turns out, when we compared citation ranks and political party we found that the published opinions of judges appointed by Republican presidents are cited significantly more often [than] those written by Democratic appointees."[174] Indeed, Choi and Gulati themselves found a similar result in later work on judicial bias.[175]

The problem of potential bias would apply equally to this study, but in many cases, there is a rather simple and elegant solution. The key issue presented by the bias problem is that the measures of ideology are not perfect, so that some of the conservative discount applied to liberals and some of the liberal discount applied to conservatives seeps into the quality ratings. This means that the quality of judges as evaluated by the opposite party is somewhat too low and that the quality of judges as evaluated by the same party is somewhat too high. This is a problem where, as the quote above asserts,[176] the ideological balance of the judiciary is lopsided. But it also means that the true quality ranking should lie somewhere between the score given by judges of the opposite party and the score given by judges of the same party. By choosing judges who are ranked relatively high and who are ranked similarly by judges of both parties, the two scores converge on one another, offering the double benefit of identifying a candidate who is not very ideological and for whom the quality rankings are free from this potential ideological contamination.[177]

---

173. Workshop on Empirical Research in the Law, *supra* note 9, at 174.

174. *Id.* at 174-75.

175. *See* Choi & Gulati, *supra* note 108; Stephen J. Choi & G. Mitu Gulati, *Ranking Judges According to Citation Bias (As a Means to Reduce Bias)*, 82 NOTRE DAME L. REV. 1279 (2007).

176. *See supra* note 173 and accompanying text.

177. Note that Choi and Gulati's method could use an analogous approach for their ranking system as well. In fact, the authors perform this analysis for Judge Posner in a follow-up piece. Choi & Gulati, *supra* note 113, at 35-36 (noting that Judge Posner is highly cited by both Republicans and Democrats).

This possibility leads to the final columns of Table II, which provide rankings for each judge by Democratic panels only (column 8) and Republican panels only (column 9). A survey of these columns reveals that in the vast majority of cases, the rankings by each party correspond to the expected ideology of the judges listed. There are a few judges whose rankings seem to be reversed, which should be expected in noisy data that results from using panels on which each judge is only one participant. Although these ideological measures seem correct in most cases, there is insufficient data to make *definitive* conclusions about particular judges' ideologies.

The technique of ranking judges separately by party is designed to avoid the problem of ideological bias and to look for judges who have high bipartisan rankings. A judge who satisfies the bipartisan criterion would have a high ranking (i.e., a low number) in both column 8 and column 9. Consider again then-Judge Sotomayor in Table II, who is the seventeenth judge listed in the table. She falls into this ideal category, receiving a high positive ranking from Democrat and Republican panels alike. Then-Judge Sotomayor, a Democratic nominee to the Second Circuit, is ranked thirty-first by Democratic panels and twenty-fifth by Republican panels – nearly identical, high ratings. Thus, not only do we know that Justice Sotomayor was highly respected during her time on the Second Circuit, we also know that this high ranking is not the result of ideological distortion.

In contrast, consider two judges at the opposite extreme, one from each political party. Judge Stephen Reinhardt, nominated by Democrat Jimmy Carter, received a score of 299 from Democratic panels and a score of 380 from Republican panels. Judge Andrew Kleinfeld, nominated by Republican George H.W. Bush, has almost the same discrepancy (with the opposite ideological interpretation), receiving a score of 381 from Democratic panels and a score of 294 from Republican panels. Indeed, Judge Kleinfeld has been identified by Professor Erwin Chemerinsky as something of a conservative counterpart of Judge Reinhardt, a hypothesis this study seems to confirm.[178] These three judges illustrate a proposition pointed out above – that judges with large ideological differentials (Reinhardt and Kleinfeld) have difficulty climbing into the higher rankings, while those with small ideological differentials (Sotomayor) have the potential to reach the top of the list.

The example of Judge Reinhardt is a telling one, as he tends to stand out in one way or another in most rankings. Cross and Lindquist, for example, found in their reversal rate ranking that Judge Reinhardt had the worst rating and interpreted that as a reason to be cautious about their ranking.[179] They

178. Chemerinsky, *supra* note 27, at 2-3 ("Conservative judges such as Diarmuid O'Scannlain and Andrew Kleinfeld are just as conservative as [J]udges Reinhardt and Pregerson are liberal."). Although the data do seem to suggest that Judge Kleinfeld is the Republican Judge Reinhardt, Judge O'Scannlain appears to be less ideological and of higher quality than does Judge Pregerson, possibly undermining Chemerinsky's comparison somewhat.

179. Cross & Lindquist, *supra* note 18, at 1414.

wrote that "we suspect that few would regard [Judge Reinhardt] as clearly the worst circuit court judge sitting on the bench" and noted that "[h]is outside circuit citations in the Choi and Gulati study exceeded the median."[180] But, as noted above, the real reason for being cautious about reversal rate rankings is the tiny sample size, not that the outcomes do not correspond to what most academics think about judges.[181] However, it is noteworthy that when Cross and Lindquist conduct their alternative "clustering" evaluation of judges, Judge Reinhardt still finds himself alone in a cluster.[182]

This study confirms that Judge Reinhardt is indeed among the most negatively cited judges on the federal appellate bench, a distinction he probably would not be terribly surprised or offended by. Judge Reinhardt's unabashed emphasis on doing justice as he sees it is in some ways analogous to Judge Posner's unabashed emphasis on pragmatic results or economic efficiency. Just as Judge Posner's law and economics approach to the law does not appeal to all judges, neither does Judge Reinhardt's freewheeling approach to the law appeal to all judges. Both are well known and highly cited among academics, but neither has embraced traditional legal reasoning as his methodological approach in deciding cases.

2. Ninth Circuit Reforms

The techniques developed in this Article could also shed light on broad questions of reform of judicial administration, such as the many proposals to split or otherwise reorganize the Ninth Circuit that have surfaced over many years.[183] In 1997, Congress created a commission to investigate restructuring the boundaries and administration of the Courts of Appeals, including a possible split of the Ninth Circuit.[184] The Commission recommended against a split but did recommend a restructuring into regional "adjudicative divisions."[185] Central to the debate was whether the Ninth Circuit had "grown to a point that it [could not] function effectively."[186] Of course, this criterion of effective functioning presupposes some measure of what "functioning effectively" entails, the development of which is one of the principal goals of this paper.

There are many types of quantitative data that one can use to ascertain how well a circuit functions. For example, the Commission looked at data such as the number of dispositions relative to cases filed, the number of oral arguments, the time from filing to disposition, and reliance on visiting judges

180. *Id.*

181. *See supra* Part II.A.

182. Cross & Lindquist, *supra* note 18, at 1428.

183. COMM'N ON STRUCTURAL ALTERNATIVES FOR THE FED. COURTS OF APPEALS, FINAL REPORT 33-34 (1998).

184. *Id.* at 1.

185. *Id.* at iii.

186. *Id.* at ix.

from other circuits.[187] This type of information is certainly relevant in evaluating the productivity or efficiency in the circuit, but information of this type does not explain anything about the quality of the circuit's products – the decisions and opinions generated by the circuit. Indeed, the quality and uniformity of the circuit's legal decisions, which translate into the "consistency and predictability of the law" in the Commission's language, are very difficult to evaluate using traditional approaches, and indeed the Commission did not have enough time to conduct a "statistically meaningful analysis."[188] One oft-recited statistic that is designed to address this issue is the rate at which the Ninth Circuit is reversed by the Supreme Court, but, as many commentators have pointed out, the reversal rate by itself may tell very little about the quality of the circuit, because the Supreme Court itself is arguably ideologically biased.[189] Thus, one is left without an objective means to evaluate the functioning of one circuit relative to another, at least if one is unwilling to equate quantity of decisions with quality.

This Article's analysis is well suited to addressing these types of questions, and the results suggest the Ninth Circuit may indeed have some serious problems. The Ninth Circuit is the most negatively cited circuit in the United States, in terms of the ranking of its judges by outside-circuit citations, as illustrated in Table VIII below.

Table VIII. Judge Quality by Circuit: Outside-Circuit Citations Only

Circuit	Median Judge Quality	Mean Judge Quality	Number of Sitting (Active and Senior) Judges*
First	0.036	0.045	10
Second	0.032	0.034	22
Third	0.030	0.034	21
Fourth	0.034	0.032	14
Fifth	0.027	0.029	22
Sixth	0.031	0.023	29
Seventh	0.035	0.028	15
Eighth	0.035	0.036	20
Ninth	0.016	0.018	48
Tenth	0.039	0.034	22
Eleventh	0.037	0.035	17
D.C.	0.032	0.031	14

* *According to the Federal Judicial Center. Retrieved from* http://www.fjc.gov/public/home.nsf/hisj *on July 4, 2009.*

187. *Id.* at 39 n.92.

188. *Id.* at 39.

189. *See, e.g.*, Cross & Lindquist, *supra* note 18, at 1404 ("[T]he Court's frequent reversal of a particular judge's decisions might reflect nothing more than ideological incompatibility between the Court's majority and the judge or opinion under review."); Richard A. Posner, *Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality*, 29 J. LEGAL STUD. 711, 712 (2000) (arguing that "reversals by the Supreme Court often involve disagreement rather than the correction of error").

The low performance of the Ninth Circuit might be expected by those who have read the Landes et al. study, but what might not be expected is that the low performance does not seem to be ideologically based. An initial objection to this result would be that the lopsided partisan makeup of the appellate courts in favor of Republican appointees might make the perceived left-leaning Ninth Circuit look worse. In contrast, the data in this Article suggests that neither part of that statement is true. First, even if one restricts the data to citing panels with a majority of Democrats, the Ninth Circuit is still by far the worst in the proportion of negative versus positive citations. Moreover, judges appointed by both Republican and Democratic presidents, on average, do not see Ninth Circuit judges very differently; if one compares the average rankings for Ninth Circuit judges in columns 8 (ranking by Democrats) and 9 (ranking by Republicans), there is only a slight difference between the two (average rank 219 by Democrats and average rank 232 by Republicans). This gap is smaller than that of several other circuits and much smaller than that of the Third Circuit (average rank 202 by Republicans and 152 by Democrats). Finally, and contrary to what is often asserted, the low performance of the Ninth Circuit also is not solely or even mostly the result of the Democratic judges in the Circuit – the Republican judges are as equally represented in the bottom half of the Circuit as the Democratic judges.

The data reveal strong evidence that the work product of the Ninth Circuit is not well regarded by federal judges in other circuits, a result consistent with the Landes et al. study.[190] The purported left-leaning ideology of the Ninth Circuit, however, is not the explanation for this low performance, although ideological disagreement (left and right alike) might have something to do with it. As discussed in Part V.D.1 above, the Ninth Circuit is an ideologically driven circuit, which cannot bode well for producing a coherent body of case law. But another possibly related explanation is indeed the traditional one of circuit size, as suggested by Table VIII.[191] In the table, there appears to be a trend of negative performance associated with larger circuits, which is consistent with what Judge Posner and many others have argued.[192] Of course, the fact that there seems to be a relationship between circuit size and circuit quality does not mean that splitting a circuit would improve anything. But it does at least provide some leverage on the policy question of whether the Ninth Circuit needs reform.

190. Landes et al., *supra* note 7, at 303-04, 318.

191. Note that the figures given in Table VIII are for current circuit size, which may not be representative of the weighted circuit size over the period covered by this study. Ideally, the latter measure would be used.

192. Posner, *supra* note 189.

VI. CONCLUSION

The results of this study suggest that the vast majority of federal appellate judges are doing good work, or at least that most of them are seen by one another as performing consistently at approximately the same level. The data strongly suggest that there are a handful of high-performing judges with a very positively regarded work product. The data also suggest that there is a relatively small group of judges whose work product is regarded quite negatively. The data also reveal that the majority of judges are indistinguishable from one another in terms of the quality of their work product, suggesting a significant degree of uniformity. This finding is similar to how Choi and Gulati's results have been interpreted – that their "main result" is that "most federal judges perform roughly equally."[193]

The judges at the high end of the scale tend to be well regarded by colleagues appointed by presidents of both political parties. In contrast, the judges at the low end are all over the map, some of whom are negatively cited by both parties and some of whom have divergent ratings by party. This suggests that there are at least two factors at work in the mix of positive and negative citations – ideology and quality – and that both high quality and balanced ideology are necessary to reach the highest levels.[194] Thus, contrary to the claims sometimes made in political science, the courts of appeals are not solely or even primarily driven by ideology but are greatly influenced by quality.[195] However, contrary to the claims made by some legal scholars and judges,[196] judicial ideology is a significant determinant in judicial decision-making for a large number, but probably a minority, of federal appellate judges.

---

193. Bybee & Miles, *supra* note 74, at 1067-68.

194. It is also possible that quality and ideology are two alternative qualifications that presidents take into account. *See, e.g.*, David C. Vladeck, *Keeping Score: The Utility of Empirical Measurements in Judicial Selection*, 32 FLA. ST. U. L. REV. 1415, 1426 n.47 (2005) ("My intuition is that judges appointed mainly because of their partisan views maintain that approach on the bench and more reliably vote in a partisan manner, while judges selected mainly for their merit, but are viewed as ideology safe, tend to put ideology aside far more often.").

195. Judge Selya pointed out that "[i]n a ranked list, someone has got to wind up at the bottom." Selya, *supra* note 151, at 1296 (suggesting that being ranked at the bottom (which Judge Selya definitely is not) may not mean much). Although that is true, it is somewhat misleading; there is a difference between being ranked last with minute differences from the penultimate judge and being ranked last with a significant difference from the adjacent judge. The latter situation is what we have in this study. Although the middle of the pack is statistically indistinguishable, the bottom is clearly the bottom.

196. Probably the most notable judicial opponent of ideological explanations of judging is Judge Harry Edwards of the D.C. Circuit. *See* Harry T. Edwards, *Collegiality and Decision Making on the D.C. Circuit*, 84 VA. L. REV. 1335 (1998).

The observation that ideology appears to play a role in the quality and uniformity of judicial output should not obscure the relative contributions of ideology and quality in the federal judiciary, however. There is indeed an ideological effect in the decision to distinguish or follow precedents, both within and between circuits. But the ideological effect, at least as that effect is measured by political party, appears to weigh much less in the decision than the quality of the panel that produced the precedent or the judge who wrote the opinion. Thus, while ideology is a significant factor for some judges, it is not the most important factor for the majority of judges.

The measures of judicial performance developed in this Article are intended to serve as an important – perhaps even a necessary – component in evaluating appellate judges, but not as the sole criterion, whether in the context of Supreme Court nominations, judicial administration, or scholarly research on the judiciary. Indeed, the quantitative information on productivity provided by the citation-count studies, as well as more traditional statistics on caseload, backlog, and the like, provide indispensable complements to the results presented here. Although those studies can measure productivity, prestige, and even influence, they do not directly measure the average quality of judicial output. Similarly, although this study can measure average quality of judicial output, it cannot measure productivity. This study indicates that both types of measures are necessary to gain a full picture of performance in the federal judiciary.

The approach taken in this Article, like the other citation studies, is particularly important to evaluating judges because the data used – citations to judicial opinions – engage directly with the reasoning in the cases, rather than relying on outcome measures such as votes on dispositions. As Choi and Gulati point out, the use of citations rather than votes allows the analysis to penetrate inside the reasoning of and precedent created by opinions, not just the outcomes.[197] But using positive and negative citations has the additional advantage of allowing a seamless interface between judicial rankings and the creation of "good law," as opposed to expansive theory. If a case has been criticized and distinguished more than followed, the case is often described as "bad law," or at least as "against the weight of authority." The virtue of this study is that judges are evaluated using the same criteria that is used to evaluate cases as precedent.

Still, there are many aspects to good judging, and this study only measures some of them. Even those who embrace Choi and Gulati's quantitative perspective say there are "multiple types of judges with multiple types of virtues and shortcomings, with no one type clearly preferable to others."[198] This does not mean, however, that good judging cannot be quantified – the problem then is one of balancing different dimensions of judging. These

197. Choi & Gulati, *supra* note 175, at 1281, 1289.

198. Cross & Lindquist, *supra* note 18, at 1414.

various virtues and shortcomings are apparent to other judges, so why not let them do the balancing? This study attempts to do just that.