NOTES

MEASURING THE EFFECTS OF SPECIALIZATION WITH CIRCUIT SPLIT RESOLUTIONS

Eric Hansford*

The standard measure of a federal circuit court's judicial performance is its reversal rate—how often the Supreme Court reverses the circuit when reviewing decisions from that circuit. This Note introduces an alternative way to measure circuit performance: circuit split resolutions.

When circuits disagree ("split"), the Supreme Court often intervenes to resolve the issue. Even though several circuits might have taken a certain position, the Supreme Court reverses only one. Because the Court does not randomly select which circuit to review, relying on reversal rates can paint a distorted picture of the Supreme Court's rate of agreement with a given court. Instead, for each circuit split resolution, this Note tracks which circuits the Court agreed with and which it disagreed with. A higher agreement rate marks better performance.

This Note uses circuit split resolution to probe whether increasing circuit specialization affects circuit performance. The results lend preliminary but inconclusive support to the proposition that increased specialization by generalist courts improves performance.

* J.D. Candidate, Stanford Law School, 2011. Thank you to Mitch Polinsky for guiding me through writing this Note. Thanks also to Jeff Fisher, Masha Hansford, Tim Kautz, and the editors of the *Stanford Law Review* for their helpful comments, suggestions, and ideas.

INTRODUCTION

Suggestions for a specialized judiciary are a lightning rod in legal academia, apt to attract passionate defenses of the idealized "generalist judge." But the dichotomy of specialist versus generalist judges is a false one. In fact, all judges specialize to some extent thanks to variety in the subject matters of their cases. Within the regional federal courts of appeals ("circuits"), the Second Circuit is known for securities law, the Fifth for immigration law, and the D.C. for administrative law. By recognizing that judicial specialization by subject matter is not binary but instead lies along a continuum, this Note aims to test empirically the predicted benefits of specialization. In particular, this Note examines whether partial specialization improves judicial performance.

A court partially specializes when it decides more cases than other courts in a particular subject matter and gains a relative advantage in deciding that kind of case. A court partially specializes *relative* to other courts: the Second Circuit partially specializes in securities law because it handles more securities cases than the other federal courts of appeals, but it has not fully specialized because it also hears nonsecurities cases. Partial specialization is about load relative to other courts; full specialization, on the other hand, is about narrowed subject matter jurisdiction of a specific court. So partial specialization, unlike full specialization, requires a comparison group. Economists have long concluded that full specialization leads to better performance. This conclusion should apply to partial specialization as well. But empirically testing these benefits for courts runs into an immediate problem: how does one assess when a court performs "better"?

This Note addresses the difficulty of measuring judicial performance by exploiting a quirk in the federal appellate system that pits courts of appeals directly against each other: the Supreme Court typically decides cases in order to resolve conflicts ("splits") among the circuits. Whether the Supreme Court agrees with a circuit's position in a given conflict provides one meaningful

measure of the circuit's judicial performance. The regional circuits[1] are considered generalist courts, but partially specialize relative to each other. In this Note, I analyze which circuit "prevails" in the Supreme Court. I sort the cases by subject matter, and then track whether the Supreme Court agrees more often with a court that is partially specialized in that subject matter. A finding that the Court agrees with partially specialized courts more often would indicate that partial specialization improves judicial performance.

Part I gives the theory behind measuring partial specialization in the regional circuits and defines the variables. Part II introduces this Note's key innovation: measuring circuit performance with circuit split resolutions instead of reversal rates. Part III describes how I collected and adjusted the data and gives summary statistics, including a comparison of circuit split resolutions and reversal rates for the 2005 through 2008 Terms. Part IV presents the results, which are limited by the small data set. No single test produces a statistically significant result, but the results are consistent with partial specialization improving judicial performance. The Conclusion considers how to make circuit split resolutions more useful in future empirical research.

I. SPECIALIZATION AND THE FEDERAL COURTS OF APPEALS

All judges specialize in one sense: their only job is judging.[2] Courts also often specialize geographically.[3] But the literature on specialized courts focuses on one particular kind of specialization: specialization in subject matter. Some quibble over when to call a court "specialized" as opposed to "generalist,"[4] but I follow conventional use and define a court as "generalist" if it has jurisdiction to hear nearly all cases in nearly all subject matters (like the D.C. Circuit) and "fully specialized" if it hears cases in only a few (like the Federal Circuit).[5]

1. The regional circuits include the First through Eleventh Circuits and the D.C. Circuit. This encompasses all federal courts of appeals except for the Federal Circuit, which is considered a fully specialized court.

2. *See* Richard A. Posner, *Will the Federal Courts of Appeals Survive Until 1984? An Essay on Delegation and Specialization of the Judicial Function*, 56 S. CAL. L. REV. 761, 778 (1983).

3. *See* Lawrence Baum, *Probing the Effects of Judicial Specialization*, 58 DUKE L.J. 1667, 1671 (2009).

4. *See, e.g.*, S. Jay Plager, *The United States Courts of Appeals, the Federal Circuit, and the Non-Regional Subject Matter Concept*, 39 AM. U. L. REV. 853, 857-60 (1990); Richard L. Revesz, *Specialized Courts and the Administrative Lawmaking System*, 138 U. PA. L. REV. 1111, 1123 (1990).

5. *See, e.g.*, Edward K. Cheng, *The Myth of the Generalist Judge*, 61 STAN. L. REV. 519, 526-27 (2008) ("[T]his Article therefore uses the term 'specialized' to denote any court or judge that deviates from the generalist ideal.").

Specialization should increase both judicial *efficiency* and *expertise*.[6] Increased judicial efficiency means that opinions can be written more quickly: an experienced securities judge can write a securities opinion more easily than could an inexperienced judge because the experienced judge already knows the relevant statutes, regulations, and case law. Increased judicial expertise leads to better opinions: a judge with a deep understanding of securities law has an advantage over a judge who knows little about the subject. Efficiency goes to the quantity of opinions, while expertise goes to the quality of opinions.

This Note focuses exclusively on whether specialization improves the quality of opinions—a predicted result of expertise. It ignores the efficiency benefits of specialization, which may allow judges to decide more cases or shift their time to other tasks. Because it ignores efficiency, this analysis likely understates the benefits of specialization.

A. *Previous Work on the Effects of Judicial Specialization*

The literature on judicial specialization usually discusses specialization in the abstract.[7] Most commentators agree that specialization will improve judicial performance.[8] But they debate whether the improvements in performance will outweigh the disadvantages that may come with full specialization. The disadvantages range from judicial boredom[9] to judges overstepping their bounds and trying to legislate from the bench.[10]

---

6. If judicial specialization leads to fewer decisionmakers, specialization may also end up increasing the uniformity and consistency of decisions. But specialization itself does not drive this effect.

7. *See, e.g.*, Baum, *supra* note 3; Henry J. Friendly, *Reactions of a Lawyer-Newly Become Judge*, 71 YALE L.J. 218, 219-29 (1961); Posner, *supra* note 2; Revesz, *supra* note 4; Diane P. Wood, *Generalist Judges in a Specialized World*, 50 SMU L. REV. 1755 (1997).

8. Some predict that specialization improves judicial performance primarily in complex subject matters. *See, e.g.*, Revesz, *supra* note 4, at 1117 ("[P]roponents of specialization argue that specialized courts are more likely to make correct decisions in complex areas."); Wood, *supra* note 7, at 1766 ("[Specialization] would ensure that the adjudicators were knowledgeable in the subject matters presented to them . . . ."). Others think specialization improves judicial performance in all areas. *See, e.g.*, Rochelle Cooper Dreyfuss, *Specialized Adjudication*, 1990 BYU L. REV. 377, 378 ("Most important, the court's expertise should enable it to craft better opinions, especially in fields where a small number of cases are now distributed rather thinly among the regional courts. Since generalist judges are confronted with the specialty subject matter infrequently, they lack the motivation, experience, and time to develop an understanding of the law. They decide the occasional case based upon a cursory understanding of policy and receive limited feedback on how well they fared. Thus, [a] specialized court's sustained involvement with a field would facilitate superior decisionmaking."). *But see* Posner, *supra* note 2, at 780 ("A person who does only one job may perform better than an abler person who divides his time among several jobs, none of which he learns to do really well. But I wonder how transferrable this insight is from the industrial, technical, and academic fields where it is conventionally articulated to appellate judging.").

9. *See* Posner, *supra* note 2, at 779.

10. *See id.* at 784-85.

Determining the actual effects of specialization requires empirical evidence.[11] Some studies have tried to find the evidence by focusing on a single fully specialized court (often the Federal Circuit),[12] or a group of fully specialized courts.[13] But for most of the fully specialized courts, "a good point of reference for comparison does not exist";[14] typically no other court decides cases involving the subject matter.[15] Even aside from the dearth of comparison groups, measuring judicial performance is tricky, and no study convincingly resolves whether specialization improves judicial performance. As Lawrence Baum explains: "[S]pecialization of judges might have significant effects on their work. Debates over specific proposals for specialized courts and the general movement toward greater specialization in the courts reflect a belief that such effects exist. The empirical evidence on the impact of specialization, however, is limited."[16]

This Note looks to the federal courts of appeals to provide some of this missing evidence. By examining partially specialized courts rather than fully specialized ones, this Note can control for many of the factors that might otherwise hinder identification of a relationship. More importantly, partially specialized courts have a clear comparison group, which permits the identification of differences in judicial performance.

B. *Specialization in the Circuits*

The United States has thirteen courts of appeals: the eleven numbered regional circuits (the First Circuit through the Eleventh Circuit), the D.C. Circuit, and the Federal Circuit. The numbered circuits and the D.C. Circuit have jurisdiction over appeals from almost all of the decisions of the federal district courts they contain,[17] so they are generalist courts.[18] The Federal Circuit,

11. *See* Baum, *supra* note 3, at 1681.

12. *See, e.g.*, Lawrence Baum, *Specialization and Authority Acceptance: The Supreme Court and Lower Federal Courts*, 47 POL. RES. Q. 693 (1994); Glynn S. Lunney, Jr., *Patent Law, the Federal Circuit, and the Supreme Court: A Quiet Revolution*, 11 SUP. CT. ECON. REV. 1 (2003); R. Polk Wagner & Lee Petherbridge, *Is the Federal Circuit Succeeding? An Empirical Assessment of Judicial Performance*, 152 U. PA. L. REV. 1105 (2004); *see also* Robert M. Howard, *Comparing the Decision Making of Specialized Courts and General Courts: An Exploration of Tax Decisions*, 26 JUST. SYS. J. 135 (2005) (discussing the U.S. Tax Court).

13. *See* Lawrence Baum, *Specializing the Federal Courts: Neutral Reforms or Efforts to Shape Judicial Policy?*, 74 JUDICATURE 217 (1991); Dreyfuss, *supra* note 8.

14. Baum, *supra* note 3, at 1681.

15. *See id.* at 1682 ("Even when generalist and specialized courts decide the same types of cases, efforts to compare them can run into measurement problems."); Ellen R. Jordan, *Specialized Courts: A Choice?*, 76 NW. U. L. REV. 745, 784 (1981).

16. Baum, *supra* note 3, at 1681.

17. *See* 28 U.S.C. §§ 1291, 1294 (2006).

18. Technically they are not true "generalist" courts because the federal courts have limited subject matter jurisdiction. *See* ERWIN CHEMERINSKY, FEDERAL JURISDICTION 259

which hears cases involving only certain subject matters, is the only specialist federal court of appeals.[19] The Federal Circuit rarely opposes other circuits in circuit splits before the Supreme Court, so this study excludes it.

While the regional courts of appeals are generalists, the subject matters of their cases vary based on the kinds of cases that arise in their districts. The D.C. Circuit presents the most extreme example, its caseload crowded with administrative law cases.[20] Similarly the Second Circuit (containing New York) has more than one-third of the federal appeals on securities law,[21] and the Fifth Circuit (containing Texas) has about half of the federal immigration law appeals,[22] although neither has a docket dominated by these cases. These levels of specialization are the most extreme.[23] But in any given subject, some circuits partially specialize relative to the other circuits.

Although at least one commentator has declared that studies of judicial specialization should focus on the judge, not the court,[24] here the most logical unit of analysis is the court. Courts of appeals make decisions in panels of

(4th ed. 2003) ("State judiciaries have general jurisdiction and may therefore hear all causes of action unless there is a statute denying them subject matter jurisdiction. But federal courts have limited subject matter jurisdiction; that is, they are restricted in what cases they may adjudicate and may exercise jurisdiction only if it is specifically authorized."). But colloquially, most consider the federal courts of appeals (other than the Federal Circuit) generalists, given that they have authority over diverse subject matters. *See, e.g.*, Ruth Bader Ginsburg, *An Overview of Court Review for Constitutionality in the United States*, 57 LA. L. REV. 1019, 1021 (1997) ("With some notable exceptions . . . federal courts are not specialized tribunals; typically, they are generalist courts, and none of their members sit, as continental judges do, in sections divided by subject matter.").

19. *See* 28 U.S.C. §§ 1291, 1295.

20. Chief Justice Roberts wrote an article highlighting the circuit's unique caseload. *See* John G. Roberts, Jr., *What Makes the D.C. Circuit Different? A Historical View*, 92 VA. L. REV. 375 (2006). For example, because the District of Columbia has no federal prisons in its borders, prisoner petitions, "which make up a notable portion of the docket nationwide on other courts of appeals," represent only a small part of the D.C. Circuit's work. *Id.* at 376. On the other hand, "about two-thirds of the cases before the D.C. Circuit involve the federal government in some civil capacity, while that figure is less than twenty-five percent nationwide." *Id.* at 377. The D.C. Circuit also has exclusive jurisdiction over certain kinds of appeals. *See, e.g.*, 26 U.S.C. § 7482(b)(3) ("In the case of any decision of the Tax Court in a proceeding under section 7478, such decision may only be reviewed by the Court of Appeals for the District of Columbia.").

The D.C. Circuit was initially included in the data collected but had to be dropped because its mix of cases diverged so significantly from that of any other generalist circuit. *See* discussion *infra* Part III.A.

21. *See* ADMIN. OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS: MARCH 31, 2009, at 32 tbl.B-7 (2009), *available at* http://www.uscourts.gov/uscourts/Statistics/FederalJudicialCaseloadStatistics/2009/tables/B07Mar09.pdf.

22. *See id.*

23. And slightly misleading since they do not correct for the size of the circuit.

24. *See* Baum, *supra* note 3, at 1672 ("I refer to 'judges' rather than 'courts' because the judge is the appropriate unit to consider. To the extent that specialization by case type affects what courts do, it is primarily because individual judges do work that has only a limited range in its subject matter.").

three. What matters is not the expertise of the opinion writer, but of the collective panel.[25] On the panel, a single knowledgeable judge can make a convincing case to her two colleagues. Further, a circuit as a whole approves each disposition, adopting the three-judge panel decision by denying the inevitable motions for a rehearing en banc and by keeping the decision alive in subsequent opinions. The court—not the judge—makes the final decision, and so the partial specialization of the court—not the judge—should be examined.

In this Note, two similar variables measure partial specialization of circuits: Yearly Cases per Judgeship and Partially Specialized. Yearly Cases per Judgeship is the number of cases a circuit has in a given subject matter,[26] controlling for circuit size.[27] The Appendix describes the precise construction of the variable. For two reasons, I measure specialization with the number of cases in a subject matter instead of the percentage of caseload involving that subject matter. First, and most importantly, economists believe that the benefits of specialization accrue through the learning that occurs by doing and by attempting to solve a problem.[28] Learning occurs with each case decided and each opinion written. Second, allocating a high percentage of the docket to a single subject matter may reflect an abnormally low overall caseload instead of an abnormally high number of cases in the subject. By focusing on the number of cases, any benefits identified are attributable to partial specialization, not a light caseload.

Partially Specialized is an indicator variable that separates the circuits with the three highest Yearly Cases per Judgeship in a subject matter from the nine lowest circuits.[29] Setting the dividing point at the top three is of course somewhat arbitrary,[30] but it was the most natural breaking point in the data, almost

25. *See* Cheng, *supra* note 5, at 547 ("Using opinion writing as a measure of specialization tendencies is a reasonable choice, but it cannot capture all of the underlying behavior. For example, the metric necessarily misses the influence that a nonwriting expert might have on the ultimate opinion, whether at conference or during the opinion writing process.").

26. The subject categories used to construct Yearly Cases per Judgeship come from coding used by the Administrative Office of the Courts. *See infra* notes 111-15.

27. Courts of appeals sit in panels of three, so an average judge in the circuit will decide three times the cases reflected by the variable.

28. *See, e.g.*, Kenneth J. Arrow, *The Economic Implications of Learning by Doing*, 29 REV. ECON. STUD. 155, 155-56 (1962).

29. The final analysis excluded the D.C. Circuit, *see infra* Part III.A, so Partially Specialized became the three highest and eight lowest.

30. I also tried dividing the circuits into the single highest and eleven lowest, the two highest and ten lowest, and the four highest and eight lowest. The results were consistent with those in Part IV and followed a predictable pattern: partial specialization had a stronger effect when designating the single top or top two circuits as "partially specialized," and a weaker effect when designating the top four circuits as "partially specialized." The difference was statistically significant when dividing the single top circuit from the other eleven circuits. Across the data, though, the most natural breaking point fell after the top three, so I used that division.

never separating circuits with similar levels of specialization.[31] Throughout this Note, I refer to a court as "partially specialized" if it numbers among the three highest Yearly Cases per Judgeship in the subject matter, and "nonspecialized" if not.

Yearly Cases per Judgeship and Partially Specialized offer slightly different notions of specialization. Yearly Cases per Judgeship measures specialization continuously, while Partially Specialized measures specialization categorically. Most legal work has focused on specialization categorically,[32] but the idea of specialization from economics is not so limited.[33] A priori neither notion is superior, so I include both.

Partially Specialized, unlike Yearly Cases per Judgeship, also makes combining subject matters easy. Partially Specialized has the same scale in each subject matter—all subjects have three partially specialized circuits—while Yearly Cases per Judgeship varies dramatically subject to subject. I create a standardized variable to pool Yearly Cases per Judgeship across subject matters,[34] but Partially Specialized has a natural subject-neutral interpretation without manipulation.

Given ways to measure specialization, the next puzzle piece in assessing the effect of specialization is finding a way to measure judicial performance.

II. MEASURING JUDICIAL PERFORMANCE WITH CIRCUIT SPLITS

When the Supreme Court takes a case from a court of appeals, it typically takes it in order to settle a disagreement among the circuits.[35] In fact the Court often declares that it granted certiorari to resolve a circuit split.[36] Splits can

31. The one exception was Labor, in which the Seventh Circuit, with 7.55 Yearly Cases per Judgeship, was designated partially specialized and the Second Circuit, with 7.13 Yearly Cases per Judgeship, was not. In the Labor circuit splits the Second Circuit performed identically to the Seventh Circuit, so the coding does not affect my results. See Table 2 for complete summary statistics.

32. *See* sources cited *supra* notes 7-16.

33. *See, e.g.*, Gary S. Becker & Kevin M. Murphy, *The Division of Labor, Coordination Costs, and Knowledge*, 107 Q.J. ECON. 1137, 1141 (1992) (considering factors that affect the "degree of specialization").

34. *See infra* note 103 and accompanying text.

35. *See* SUP. CT. R. 10(a); *see also* ROBERT L. STERN, EUGENE GRESSMAN, STEPHEN M. SHAPIRO & KENNETH S. GELLER, SUPREME COURT PRACTICE 226 (8th ed. 2002) ("The Supreme Court often, but not necessarily, will grant certiorari where the decision of a federal court of appeals, as to which review is sought, is in direct conflict with a decision of another court of appeals on the same matter of [law]. One of the prime purposes of the certiorari jurisdiction is to bring about uniformity of decisions on these matters among the federal courts of appeals. Hence a square and irreconcilable conflict of this nature ordinarily should be enough to secure review, assuming that the underlying question has substantial practical importance." (emphasis omitted) (citation omitted)).

36. *See, e.g.*, Forest Grove Sch. Dist. v. T.A., 129 S. Ct. 2484, 2490 (2009) ("Because the Courts of Appeals that have considered this question have reached inconsistent results,

range from two courts disagreeing[37] to almost all of the circuits divided between two opposing sides.[38] To measure judicial performance, this Note uses the Supreme Court's resolution of circuit splits. I call the Supreme Court's rate of agreement with a given circuit in resolving the splits the circuit's "agreement rate."

A. *Judicial Performance and the Federal Courts of Appeals*

"Judicial performance" defies easy definition. A recent symposium on judicial performance[39] failed to provide a consensus description. Many proposals offer characteristics of good judges, only some of which translate into characteristics of good courts. Still, certain characteristics surface again and again in the literature. One influential list focuses on the "[q]uality of the [j]udicial [p]roduct."[40] Another argues that academics generally agree that judges must aim for "[j]ustice as lawfulness," meaning judges should rely not on "private judgments about fairness" but on the "public judgments" manifested in the law.[41] According to now-Justice Ginsburg the "core values by which court performance may be measured" are "fairness" and "[g]etting it right."[42]

Accepting that judicial performance includes getting it right, "[t]here is no objective external test for the 'legal correctness' of a decision or the 'legal quality' of an opinion"[43]:

---

we granted certiorari to determine whether § 1412(a)(10)(C) establishes a categorical bar to tuition reimbursement for students who have not previously received special-education services under the authority of a public education agency." (footnote omitted)); Bridge v. Phoenix Bond & Indem. Co., 553 U.S. 639, 646 (2008) ("We granted certiorari to resolve the conflict among the Courts of Appeals on 'the substantial question' whether first-party reliance is an element of a civil RICO claim predicated on mail fraud." (citations omitted) (quoting Anza v. Ideal Steel Supply Corp., 547 U.S. 451, 461 (2006)); Logan v. United States, 552 U.S. 23, 30 (2007) ("We granted certiorari to resolve a split among the Circuits as to whether § 921(a)(20)'s exception for 'civil rights restored' should be interpreted to include civil rights retained at all times." (citation omitted)).

37. *See, e.g.*, Dean v. United States, 129 S. Ct. 1849, 1852-53 (2009) (citing two sole opposing circuits as the "conflict among the Circuits").

38. *See, e.g.*, Irizarry v. United States, 553 U.S. 708, 713 n.1 (2008) (identifying a split with five circuits on one side and six on the other).

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

40. Stephen Choi & Mitu Gulati, *A Tournament of Judges?*, 92 CALIF. L. REV. 299, 305 (2004). It also lists "[c]aseload [p]erformance" (efficiency) and "[i]ndependence" as positive characteristics of judges. *Id.* at 309-10.

41. Lawrence B. Solum, *A Tournament of Virtue*, 32 FLA. ST. U. L. REV. 1365, 1382-83 (2005).

42. Ruth Bader Ginsburg, *The Obligation to Reason Why*, 37 U. FLA. L. REV. 205, 206 (1985).

43. Frank B. Cross & Stefanie Lindquist, *Judging the Judges*, 58 DUKE L.J. 1383, 1403 (2009).

> Some will see the search for a set of objective measures as pointless because they think that there is no way to measure or quantify what it means to be a good, let alone great, judge. This is likely true as an *absolute* matter. Nonetheless . . . it may still be possible to make meaningful *relative* evaluations.[44]

One common measure of judicial performance is the rate at which the Supreme Court reverses a particular judge (or circuit) when hearing a direct appeal of that judge's (or circuit's) opinion. The general consensus seems to be that reversal rates, while "imperfect," offer a "reasonable approach to evaluating judicial quality."[45]

The Supreme Court's resolution of circuit splits, which this Note puts forward, provides a better measure of judicial quality; it eliminates many of the biases of reversal rates discussed in Part II.B, but still evaluates circuits with an objective measure of quality. To see how, consider the three plausible aims for a federal appellate court deciding a difficult case. First, a court could try to reach the "objectively correct" answer in deciding an issue.[46] Second, a court could embrace its role as a subordinate court and try to match how the Supreme Court would decide the issue. Third, a court could try to enact its own policy preferences (or, less cynically, enact the solution the judges believe leads to the best policy) within the constraints set by the Supreme Court.[47] Resolutions of circuit splits measure a circuit's success in meeting either of the first two aims, and the third aim should be repudiated.

If a circuit ought to hold the first aim of finding the "objectively correct" answer, the Supreme Court's agreement rate with the circuit will be a meaningful, although indirect, measure of the circuit's performance. Finding this answer would also be the aim of the Supreme Court, and the Supreme Court is almost certainly better at finding it: Individually the Justices are among the greatest legal thinkers in the country, and collectively more than half of their

44. Stephen J. Choi & G. Mitu Gulati, *Choosing the Next Supreme Court Justice: An Empirical Ranking of Judge Performance*, 78 S. CAL. L. REV. 23, 30 (2004).

45. Cross & Lindquist, *supra* note 43, at 1403; *see also* sources cited *infra* notes 57-63.

46. This aim seems least realistic because the hardest questions of law probably have no "objectively correct" answer. If two canons point in opposite ways for a hard question of statutory construction, it will be arbitrary to pick between the two without another guiding principle (e.g., how the Supreme Court would decide the issue, or enacting policy preferences).

47. The literature on judicial specialization supports the claim that these are the three most plausible aims. Lawrence Baum identifies two possible meanings of "[g]etting decisions right." Baum, *supra* note 3, at 1676. He says that in "most" discussions, "it implicitly refers to applying the law to the facts *properly*." *Id.* (emphasis added). Baum does not define "properly," but presumably he means either a purely objective assessment of the application in the abstract (the first aim I list) or an assessment according to a certain goal of the application (such as the goal of the second aim I list—to do what the Supreme Court would do). Baum also says "[g]etting decisions right" could mean making decisions "that best reflect[] a judge's conception of good public policy." *Id.* This is the third aim I list.

cases are resolved with one or no dissenting votes.[48] Additionally, the Supreme Court "has the advantage of having considerably more time to evaluate the legal issues."[49] And "[t]he Court often has far better legal and other information on which to ground its decisions."[50]

The second aim, matching the Supreme Court, is more plausible than it might sound. After all, "the Court's decisions *are* the law of the land,"[51] so in one sense the Court always gets the answer right. The circuits are subordinate to the Supreme Court and have an obligation to follow the Court's precedent, so the view of the Supreme Court should at least guide any circuit decision.[52] If a circuit should aim to match the Supreme Court, the Court's agreement rate with the circuit directly measures the circuit's performance.

But if a circuit ought to hold the third aim of enacting policy preferences, the agreement rate will indicate nothing about the performance of the circuit. Indeed, nothing will measure how well a circuit enacts its policy preferences—it depends solely on the individual judges' views. The permissibility of holding the third aim alone, however, is questionable. A judge is not a policymaker. Most agree that judges must put aside their own private views.[53] More concretely, many commentators view the reversal rate of a court as a convincing measure of judicial performance.[54]

Thus, whether the Supreme Court agrees with a circuit court provides a reasonable measure of judicial performance. If the Supreme Court agrees more often with partially specialized circuit courts, this indicates that partially specialized courts perform better.

* * *

If Supreme Court agreement rate increases with increasing partial specialization, it may be tempting to point out that correlation does not entail causation; causation could run either way (i.e., the Supreme Court could be deferring to the more specialized courts). But "downward deference" is unlikely in the context of circuit splits for several reasons. Most importantly, splits do not systematically pit the most partially specialized circuits against the least, so it is unclear how the Supreme Court could defer downward when partially specia-

48. *See Decisions by Vote Split*, SCOTUSBLOG (July 7, 2010), http://www.scotusblog.com/wp-content/uploads/2010/07/Final-Charts-070710-41.pdf.

49. Cross & Lindquist, *supra* note 43, at 1403.

50. *Id.*

51. *Id.*

52. *See id.* ("[C]ircuit courts may be regarded as agents of the Supreme Court, so it seems appropriate to consider the evaluations of their principal." (footnote omitted)).

53. *See, e.g.*, Solum, *supra* note 41, at 1382-83.

54. *See infra* text accompanying notes 57-63.

lized courts do not adopt a single side.[55] Further, the idea that the Supreme Court would schedule full argument and write full opinions simply to defer downward seems implausible; it would be a monumentally wasteful way of making courts follow the expert circuit. (If the Court knows which position it will adopt, it will probably dispose of the case summarily and instruct circuits to follow a given decision.) Finally, few Justices will want to put aside their own judgments and defer to the lower courts they supervise.

No other obvious factors independently and systematically link agreement rates and partial specialization.[56] Therefore, if this Note identifies a positive relationship between Supreme Court agreement rate and specialization, increased partial specialization probably *causes* improved judicial performance.

B. *Measuring Judicial Performance with Circuit Split Resolutions, Not Reversal Rates*

Most other measurements of judicial performance of the federal courts of appeals (as opposed to individual judges) use the rate of reversal by the Supreme Court. A circuit's reversal rate is the number of its decisions that the Supreme Court reverses divided by the total number of its decisions that the Supreme Court reviews on the merits. Leading Supreme Court commentators,[57] the popular press,[58] and legal journals[59] use reversal rates.

---

55. The downward deference story may seem more compelling in the few cases of a dominant "expert" circuit, but not many examples come to mind beyond the Second Circuit in securities law, the D.C. Circuit in administrative law, and the Fifth Circuit in immigration law. Of these three examples, only the Second Circuit in securities law is included in my sample (immigration law is criminal, and the D.C. Circuit was excluded). In the one Second Circuit case in a securities circuit split, the Court indeed agreed with the Second Circuit. But if some circuits really are perceived as such "experts" that they receive downward deference from the Supreme Court, it seems that the other circuits would defer to them as well (or reverse positions to conform to the "expert" circuit's opinion), obviating the need for the Supreme Court to resolve a circuit split in the first place.

56. The quality of the local bar in a given subject matter could link agreement rates with partial specialization. Regions that partially specialize in certain subjects probably often have better legal representation, and better legal representation probably makes courts more likely to "get it right." This relationship will not always hold: while the Fifth Circuit partially specializes in immigration law, *see supra* text accompanying note 22, its immigration lawyers are not particularly celebrated. But it likely does hold for the securities lawyers in the Second Circuit and the administrative lawyers in the D.C. Circuit, the two other cases of extreme specialization. *See supra* text accompanying notes 20-23. For the subjects studied in this Note, *see infra* Figure 1, securities law seems to be the only one with dramatically different lawyer quality circuit to circuit. While it appears unlikely that the quality of representation drives the results, I cannot disprove that it might have some influence.

57. *See, e.g.*, *Circuit Scorecard*, SCOTUSBLOG (July 7, 2010), http://www.scotusblog.com/wp-content/uploads/2010/07/Final-Charts-070710-10.pdf (using reversal rates as the unit of measurement in a "Circuit Scorecard").

58. *See, e.g.*, Carol J. Williams, *U.S. Supreme Court Looks over 9th Circuit's Shoulder*, L.A. TIMES, June 29, 2009, http://articles.latimes.com/print/2009/jun/29/local/me-9th-scotus29 (attempting to explain why the Ninth Circuit has a higher reversal rate than aver-

Within academia, measures of circuit performance most often appear in discussions about the Ninth Circuit. The Ninth Circuit has long been the most reversed circuit,[60] and the scholarship examining why this is takes reversal rates as an important measure of performance.[61] Some academics, such as Judge Posner, propound a more nuanced version of reversal rates: the rate of unanimous (or nearly unanimous) reversals.[62] Judge Posner says these reversals indicate that the lower court decision is "more likely to be just plain incorrect, rather than merely the reflection of political difference."[63] But all versions of reversal rates depend on which decisions and circuits the Supreme Court chooses to review.

Reversal rates imperfectly measure judicial performance, however, because they pick up other effects unrelated to the quality of a circuit's decision. Circuits often act identically by choosing the same side of a circuit split, yet the Supreme Court reverses only one. The Court does not randomly select which circuit to review, distorting conclusions founded on reversal rates. Throughout this Subpart, I invoke the example of the Ninth Circuit to illustrate the biases, but the analysis applies equally to the other circuits.

The reversal rate of a circuit does not necessarily reflect the rate at which the Supreme Court disagrees with the circuit. Judge Jerome Farris of the Ninth Circuit offers the example of *California Division of Labor Standards Enforcement v. Dillingham Construction*,[64] a unanimous reversal of a Ninth Circuit decision by the Supreme Court.[65] The Supreme Court's decision in *Dillingham Construction* increased the Ninth Circuit's rate of reversal and rate of unanimous reversal. As Judge Farris points out,[66] however, the Ninth Circuit decision in *Dillingham Construction* followed a Tenth Circuit decision.[67] While the reversal counts as a black mark for the Ninth Circuit, the Tenth Circuit, which decided the same issue the same way, escapes unscathed.

age); Dan Levine, *9th Circuit Racks Up Usual High Reversal Rate in Supreme Court Term*, LAW.COM (June 30, 2008), http://www.law.com/jsp/article.jsp?id=1202422620128.

59. *See, e.g.*, Roy E. Hofer, *Supreme Court Reversal Rates: Evaluating the Federal Courts of Appeals*, LANDSLIDE, Jan.-Feb. 2010, at 8. The Hofer article even assigns each circuit a letter grade (from A to F) based on its reversal rate. *See id.*

60. *See* Kevin M. Scott, *Supreme Court Reversals of the Ninth Circuit*, 48 ARIZ. L. REV. 341, 341-43 (2006).

61. *See, e.g.*, Jerome Farris, *The Ninth Circuit—Most Maligned Circuit in the Country—Fact or Fiction?*, 58 OHIO ST. L.J. 1465 (1997); Richard A. Posner, *Is the Ninth Circuit Too Large? A Statistical Study of Judicial Quality*, 29 J. LEGAL STUD. 711 (2000); Scott, *supra* note 60; *see also, e.g.*, Cross & Lindquist, *supra* note 43, at 1402-14 (measuring judicial performance of individual judges with reversal rates).

62. *See, e.g.*, Posner, *supra* note 61, at 716 tbl.4; Scott, *supra* note 60, at 342.

63. Posner, *supra* note 61, at 716.

64. 519 U.S. 316 (1997).

65. *See* Farris, *supra* note 61, at 1466.

66. *See id.*

67. *See* 519 U.S. at 323.

If the Supreme Court randomly chose which circuit on a given side of a split to reverse, the disparate treatment of the Ninth and Tenth Circuits would be unproblematic; by the law of large numbers, the observed reversal rates should converge to each circuit's "true" reversal rate. But there is reason to think that the Court is biased in selecting which circuit to reverse.

Most concretely, a recent study finds that Justices who previously served on a circuit court appear biased when dealing with their former court.[68] The study finds statistically significant bias by Justices Ginsburg, Thomas, and Scalia in favor of the D.C. Circuit, by Justice Breyer in favor of the First Circuit, by Justice Stevens in favor of the Seventh Circuit, and by Justice Kennedy *against* the Ninth Circuit. According to the study, among the Justices serving between the 2005 and 2008 Terms, only Justice Souter appeared unbiased with respect to his former court (the First Circuit). (Chief Justice Roberts and Justice Alito were excluded because of insufficient data.)[69] Surely any change in reversal rates from home-circuit bias is small,[70] but finding such clear bias for reasons unrelated to judicial performance buttresses the case against reversal rates.

Second, circumstantial evidence indicates that the Supreme Court may seek to reverse certain judges and circuits when given the chance, which would create bias. One article quotes law professors and former Supreme Court clerks who say that the Court tries to reverse certain liberal judges (like Judge Stephen Reinhardt) and police the perceived tendency of the Ninth Circuit to rule for the "poor and powerless."[71] In support of this claim, a recent study finds that Judge Reinhardt is by far the most frequently reversed appellate judge in the country.[72] But dislike for certain segments of the Ninth Circuit does not entail consistent disagreement with the Ninth Circuit. Instead, when the Court plans to resolve a split against the Ninth Circuit, the Court can choose to use the Ninth Circuit and its liberal judges as the vehicle for reversal. (Similarly, the Court can avoid using Ninth Circuit opinions as a vehicle for affirmance.) These practices would give the Ninth Circuit a reversal rate far higher than the Court's true rate of disagreement.

68. *See* Lee Epstein et al., *Circuit Effects: How the Norm of Federal Judicial Experience Biases the Supreme Court*, 157 U. PA. L. REV. 833, 873-77 (2009).

69. *See id.* at 874 fig.9.

70. Although all the home-circuit biases that the study identifies work against the Ninth Circuit.

71. Williams, *supra* note 58; *see also* John Schwartz, *'Liberal' Reputation Precedes Ninth Circuit Court*, N.Y. TIMES, Apr. 25, 2010, at A33 ("The highest court in the land, [Supreme Court practitioner Tom Goldstein] suggested, so clearly rankles at the views of Judge Reinhardt that litigators hoping to get his decisions overturned will go out of their way to cite him by name in their briefs.").

72. *See* Cross & Lindquist, *supra* note 43, at 1407 tbl.1 (showing that, between 1989 and 2000, Judge Reinhardt was reversed fourteen times, while the next most reversed judge was reversed only nine times).

Third, certain circuits may be reversed more because of timing. The Supreme Court reverses far more often than it affirms.[73] It typically (but not always) takes cases from the federal courts to settle circuit splits.[74] A split demands circuits on both sides. Circuits that consistently weigh in on issues before others would be reversed in splits less often because the split would not yet have developed; circuits that weigh in later would be reversed in splits more often. A judicial performance measure that aggregates split and nonsplit cases should show systematically higher reversal rates for circuits that tend to rule later. In fact the circuits vary in how long they take between the district court filing and the final circuit decision[75]: in 2009, the difference between the quickest circuit (the Fifth) and the slowest circuit (the Second) was almost two years.[76] The Ninth Circuit was the second slowest.[77]

Finally, and counterintuitively, a high reversal rate may indicate better circuit performance, at least compared with the other circuits on the same side of the split. The Supreme Court chooses whether and when to decide a legal issue. Circuit biases aside, the Justices will prefer to deal with a lower court opinion that squarely frames the issue of concern and contains few peripheral issues to cloud the analysis.[78] Although this depends in part on the facts of the case, circuit courts affect it as well. An opinion that cleanly separates the various issues is ideally suited for review; an opinion that jumbles the issues together, using arguments in the alternative ("even if we agreed with the petitioner on issue *X*, she might still lose because of issue *Y*"), is not. Circuits likely to write "clean" opinions, normally regarded as a positive judicial characteristic, may be systematically picked out for review when the Court resolves a circuit split.

Outside of the context of circuit splits, higher reversal rates need not indicate inferior judicial performance. The difficulty of cases (and especially the number of exceptionally complex cases) will vary across circuits, and a harder case is more likely to be reversed. But the higher reversal rate reflects caseload, not performance. The Ninth Circuit again helps illustrate: The Ninth Circuit

73. *See Circuit Scorecard*, *supra* note 57 (showing more than 70% of Supreme Court opinions in the 2009 Term were reversals).

74. *See supra* note 35 and accompanying text.

75. *See* ADMIN. OFFICE OF THE U.S. COURTS, JUDICIAL BUSINESS OF THE UNITED STATES COURTS: 2009 ANNUAL REPORT OF THE DIRECTOR 103 tbl.B-4 (2010), *available at* http://www.uscourts.gov/uscourts/Statistics/JudicialBusiness/2009/JudicialBusinespdfversion.pdf; *see also id.* at 104 tbl.B-4A (showing the circuit-by-circuit delays at each stage of the process broken down by type of case).

76. *Id.* at 103 tbl.B-4. The average circuit took about thirty-two months.

77. *Id.*

78. *See* STERN, GRESSMAN, SHAPIRO & GELLER, *supra* note 35, at 231 ("If it appears that upon a grant of certiorari the Supreme Court might be able to decide the case on another ground and thus not reach the point upon which there is a conflict, the conflict itself may not be a sufficient reason for granting review. . . . 'Resolution here of the . . . [issue in conflict among the circuits] can await a day when the issue is posed less abstractly.'" (second omission and alteration in original) (quoting The Monrosa v. Carbon Black Exp., Inc., 359 U.S. 180, 184 (1959))).

contains the West Coast, an area uniquely concentrated in high-tech industries. These industries force the Ninth Circuit to resolve novel issues that other circuits avoid.[79] The Ninth Circuit also contains states like California apt to pass laws that push constitutional limits. Initiative and proposition processes abound in the West and often lead to innovative laws.[80] For example, the only current federal constitutional challenge to gay marriage bans is in California, following Proposition 8,[81] even though many states ban gay marriage. The Ninth Circuit will likely be the only circuit forced to deal with the thorny gay marriage issue before Supreme Court review. But reversal rates fail to control for the variation in case difficulty.

* * *

By using circuit split resolutions instead of reversal rates, this Note controls for possible biases of the Supreme Court in deciding which circuit to reverse.[82] Circuit split resolutions treat all circuits in a split equally, so no circuit should be systematically prejudiced. Further, circuit split resolutions generate more data than do reversals; each time the Court resolves a circuit split, it implicitly evaluates the decisions of the other circuits involved in the split. The circuit split data that I collect indeed paints a different picture of circuit performance, indicating that worries about bias in reversal rates are well founded. When measured by circuit split resolutions the Ninth Circuit was average.[83]

Circuit splits almost by definition involve difficult legal issues. For easy issues the circuits will probably agree. This helps judicial evaluation: courts differentiate themselves in answering difficult questions of law, not easy ones.

Circuit splits may appear to have two drawbacks, but upon close review neither undermines efforts to measure judicial performance. First, circuit splits are biased in measuring the "true" rate at which a court "correctly" decides cases of a certain difficulty, although they reflect the relative performance of courts.[84] This bias raises few practical problems—measurements of judicial

79. *See* Williams, *supra* note 58 ("[The Ninth Circuit] produces a lot of cutting-edge law, due to industries concentrated in the circuit and the large variation of underlying states and state criminal laws." (quoting Professor David Hoffman)).

80. *See id.* ("A lot of important policy cases involving interesting and difficult questions come out of the 9th Circuit. The West is known for its experimentation, the initiative process—things that bring constitutional questions to the fore more often." (quoting Professor Jeffrey L. Fisher)).

81. *See* Jesse McKinley, Bush v. Gore *Foes Join to Fight Gay Marriage Ban*, N.Y. TIMES, May 28, 2009, at A1.

82. Not every factor would cause statistical bias in identifying a relationship between judicial performance and circuit partial specialization. But some would, and all would make the identification more difficult by adding noise.

83. *See infra* Table 1.

84. This arises because the Supreme Court's agreement rate in circuit splits with a given circuit reflects only instances when the circuit has disagreed with another circuit. For ex-

performance aim at relative performance, not absolute performance.[85] Further, the bias quickly decreases as the number of courts that could disagree increases.[86] With twelve circuits, any bias in measuring the "true" rate will be trivial.

Second, circuit split resolutions leave out Supreme Court decisions that do not resolve a split, even though every decision passes some judgment on the circuit from which it arises. But Supreme Court review of circuit decisions outside the context of splits almost certainly reflects a novel legal issue.[87] Measuring the difficulty of the issue, as required to translate the Supreme Court's resolution into a measure of judicial performance, presents a nearly impossible task. Circuit splits at least provide a reliable and objective measure of judicial performance unlikely to be systematically distorted.

III. DATA COLLECTION AND SUMMARY STATISTICS

A. *Data Collection and Adjustments*

Because no one had systematically examined Supreme Court resolutions of circuit splits, all of the underlying data had to be collected manually. I first tried to identify all of the circuit splits from four recent Terms (2005 through 2008).[88] I limited myself to instances when the Supreme Court announces it is

ample, imagine a "good" court that decides questions of a given difficulty correctly 90% of the time, and a "bad" court that decides the questions correctly 60% of the time (little better than a coin flip). These are the "true rates." Both courts will decide the question correctly 54% of the time. Both courts will decide the question wrongly 4% of the time. The courts will disagree the other 42% of the time: 36% of the time the good court will get it right while the bad court will get it wrong, and 6% of the time the good court will get it wrong and the bad court will get it right. Within the instances the courts split, the good court will appear to be right 86% of the time (36/42) and the bad court will appear to be right 14% of the time (6/42). The good court's observed rate is only slightly lower than its "true rate," while the bad court's observed rate is much lower.

85. *See supra* text accompanying note 44.

86. Continuing the example from note 84, imagine three good courts and three bad courts. The courts will all decide the question correctly about 16% of the time and all decide the case wrongly a trivial number of times. The other roughly 84% of the time the courts will split. Within the 84% (the circuit splits), a given good court will appear right 88% of the time and a given bad court will appear right 53% of the time. So observed rates with six courts come much closer to the true rates of 90% and 60%. As the number of courts increases the observed values would continue to approach the true rates.

87. *See* STERN, GRESSMAN, SHAPIRO & GELLER, *supra* note 35, at 255 ("It has been reiterated many times that the Supreme Court is not primarily concerned with the correction of errors in lower court decisions. . . . The Court's aim, rather, is to resolve the conflicts among the lower courts and to determine questions of importance. . . . The Supreme Court's burden and responsibility are too great to permit it to review and correct every misstep made by the lower courts in the application of accepted principles.").

88. See the Appendix for how the splits were identified. These four Terms were selected because, with the exception of the very start of the 2005 Term, the Supreme Court included the same nine Justices. Consistency minimizes worries that the Court may be deciding cases differently at different times in the sample.

resolving a circuit split; going beyond these instances would have forced subjective assessments of whether the Court resolves a "true" circuit split, and where the apparent conflict is reconcilable. This method certainly leaves out some circuit splits that the Court resolved;[89] it should not, however, systematically exclude certain types of cases, so the omissions should not bias the ultimate findings.[90]

I then took all of the cases identified as part of the split—marking their circuits—and coded them according to whether the Supreme Court resolved the split by agreeing or disagreeing with them.

Separating the data by subject matter, as is necessary to measure the effects of partial specialization, forced two major adjustments. First, many circuit splits dropped out because they did not fit into a subject matter category. For a few splits this was because the provider of the data, the Administrative Office of the Courts, does not code for the relevant category; this eliminated civil RICO cases and Railroad Revitalization and Regulatory Reform Act cases. More often I dropped cases because the issue arises in multiple areas (such as rules of procedure or constitutional challenges) so that no single subject matter category could capture a circuit's workload in a given type of case. Overall this problem excluded about half of the observations.

Second, the D.C. Circuit had to be excluded. Although the D.C. Circuit is a generalist court, its caseload heavily tilts toward administrative law.[91] Because of the emphasis, the D.C. Circuit had by far the lowest average number of cases in each of the eight subject matters with resolved circuit splits and would have been an outlier in every split. (It often had about ten percent of the average cir-

89. Two opinions said that they were resolving circuit splits, but failed to lay out the relevant circuits. *See* Boyle v. United States, 129 S. Ct. 2237, 2242-43 (2009) ("Petitioner was then resentenced, and we granted certiorari to resolve conflicts among the Courts of Appeals concerning the meaning of a RICO enterprise." (citation omitted)); Zedner v. United States, 547 U.S. 489, 497 (2006) ("We granted certiorari to resolve the disagreement among the Courts of Appeals on the standard for analyzing whether a defendant has made an effective waiver of rights under the Act."). But these cases would have been dropped in later analysis because both concern criminal law.

90. As these omissions highlight, relying on the Supreme Court's identification of circuit splits is probably not the best identification method. The method relies on the Justices to be consistent in identifying the circuits in a split when resolving it. But the Justices may have incentives to present a split in a certain light, or inclinations toward not including the full split (or not including the split at all).

The better method would be to look at the legal question the Supreme Court resolved, and then to look back through the circuits (presumably relying on the petition for writ of certiorari) to find which courts had weighed in on the question. This is a time-consuming task that would take more than Westlaw searches. The splits resolved would probably be easiest to identify contemporaneously with the Supreme Court decisions, suggesting an institution like SCOTUSblog would be in the best position to collect this data.

91. *See supra* note 20 and accompanying text.

cuit's Yearly Cases per Judgeship.) This excluded seven observations that would otherwise have been included.[92]

I also omit criminal cases. Initially, I tried to include criminal cases by sorting them into offense categories; most criminal Supreme Court cases, however, do not involve specific offenses, so this sorting required dropping more than half of the criminal observations and often meant missing the hardest legal issues. The omitted cases typically involved constitutional rights that attach in all criminal cases[93] or sentencing decisions.[94] The more fundamental problem with criminal cases is that experience across offense categories may be essentially interchangeable. Difficult criminal issues are less isolated and more analogous to legal issues in other parts of the criminal law: experience interpreting the mens rea requirement in a drug possession statute helps decide a difficult question about the mens rea required in a statute penalizing firearm possession. In contrast, experience with tax cases is useless in resolving difficult legal questions about civil rights or contracts or postconviction proceedings.

The Appendix summarizes the sources of the data used in this Note and the mechanics of the data collection.

B. *Summary of Data*

Table 1 summarizes the Supreme Court agreement rates, broken down by circuit, for the circuit splits collected in the 2005 Term through the 2008 Term. The table includes observations that were omitted from the final analysis. For comparison, Table 1 also includes affirmance rates during the period. (The affirmance rate is the opposite of the reversal rate—it is the rate at which the circuit was affirmed.) The ranking of circuits by Supreme Court agreement rate bears almost no relation to the ranking of circuits by affirmance rate; in fact the rankings are effectively uncorrelated.[95]

The worries about measuring circuit performance with reversal (or affirmance) rates[96] appear well founded. For example, by affirmance rate, the Ninth Circuit is by far the worst performing circuit, half as likely to be affirmed as the third-worst performing. But measuring performance with the agreement rate makes the Ninth Circuit look utterly typical: its agreement rate is about average, and it performed better than four other circuits. This finding casts doubt on the key premise of the literature that aims to explain the underperformance of the Ninth Circuit.

92. I tried running some of the tests described in the next Part with the D.C. Circuit included, and nothing changed significantly.

93. *See, e.g.*, Arizona v. Gant, 129 S. Ct. 1710 (2009) (search incident to arrest).

94. *See, e.g.*, Kimbrough v. United States, 552 U.S. 85 (2007) (Federal Sentencing Guidelines).

95. The correlation is 0.012, meaning that a single additional reversal of the D.C. Circuit (for example) would have made the correlation negative.

96. *See supra* Part II.B.

The Supreme Court agreement rate rankings do not seem explicable on "political" grounds. The Second, Third, and Ninth Circuits, which have been identified as the most liberal,[97] ranked in the middle. The Fourth, Fifth, and Sixth Circuits, identified as the most conservative,[98] ranked second to last, first, and last. Conclusions about specialization based on Supreme Court agreement rates therefore likely reflect more than "politics."

97. *See* Lee Epstein et al., *The Judicial Common Space*, 23 J.L. ECON. & ORG. 303, 312 (2007) (Second and Ninth); Schwartz, *supra* note 71 (Second, Third, and Ninth).

98. *See* Epstein et al., *supra* note 97, at 312 (Fourth and Fifth); Schwartz, *supra* note 71 (Fifth and Sixth).

TABLE 1
Supreme Court Agreement Rate by Circuit,
2005 Term Through 2008 Term

Circuit	Observations	Observed Supreme Court Agreement Rate (Rank)	Affirmance Rate[99] (Rank)
First	27	48.1% (Tied at 9th)	50.0% (1st)
Second	35	54.3% (7th)	35.7% (4th)
Third	32	59.4% (5th)	16.7% (11th)
Fourth	34	47.1% (11th)	26.7% (Tied at 8th)
Fifth	37	64.9% (1st)	22.2% (10th)
Sixth	35	37.1% (12th)	27.3% (7th)
Seventh	42	59.5% (4th)	46.2% (3rd)
Eighth	27	48.1% (Tied at 9th)	26.7% (Tied at 8th)
Ninth	42	50.0% (8th)	10.8%[100] (12th)
Tenth	24	62.5% (2nd)	33.3% (5th)
Eleventh	29	58.6% (6th)	47.4% (2nd)
D.C.	21	61.9% (3rd)	30.0% (6th)
Overall	*385*	*54.0%*	*26.9%*

Table 2 summarizes Yearly Cases per Judgeship for the entire population of cases (not only those observed in the sample). For each subject matter it also

99. Compiled from SCOTUSblog end-of-Term "StatPacks." *See Circuit Scorecard—OT08*, SCOTUSBLOG (June 29, 2009), http://www.scotusblog.com/wp-content/uploads/2009/06/circuit3.pdf; *Circuit Scorecard—OT07*, SCOTUSBLOG, http://www.scotusblog.com/wp-content/uploads/2008/06/scorecard07.pdf (last visited Oct. 30, 2010); *Circuit Scorecard—OT06*, SCOTUSBLOG, http://www.scotusblog.com/archives/ScorecardOT06.pdf (last visited Oct. 30, 2010); *SCOTUSBlog Circuits Chart: October 2005 Term*, SCOTUSBLOG, http://www.scotusblog.com/archives/CircuitsFinal.pdf (last visited Oct. 30, 2010).

100. The Ninth Circuit was also affirmed in part and reversed in part 6.2% of the time. *See* sources cited *supra* note 99. No other court was affirmed in part and reversed in part. Giving the Ninth Circuit "half credit" for the cases affirmed in part and reversed in part would make its affirmance rate 13.9%.

includes the lowest number of cases required to be deemed "partially specialized."

TABLE 2
Summary of Yearly Cases per Judgeship for First Through Eleventh Circuits, by Subject Matter

Subject Matter	Mean	Standard Deviation	Max	Min	Partially Specialized Dividing Point
Civil Rights	36.4	11.0	62.6	25.2	43.1
Contracts	13.5	4.0	21.1	9.2	18.7
Labor	6.0	2.2	10.4	3.1	7.5
Postconviction	67.1	27.4	124.4	34.5	88.2
Prisoner Rights	30.1	11.9	52.3	10.4	39.1
Securities	1.5	1.2	4.6	0.6	1.8
Tax	1.1	0.3	1.9	0.7	1.4
Torts	9.9	3.2	16.1	6.1	12.4

Table 3 gives the summary statistics for circuit split decisions observed within the sample. After data adjustments, 156 observations remain. The third column gives the subject-by-subject average of Yearly Cases per Judgeship for circuits immersed in a circuit split that the Supreme Court resolves. The final column gives the total number of partially specialized courts in the sample for each subject matter.

TABLE 3
Summary Statistics for Observed Circuit Splits, by Subject Matter

Subject Matter	Observations	Sample Mean of Yearly Cases/Judgeship	Partially Specialized Circuits in Sample
Civil Rights	52	41.4	20
Contracts	12	14.2	3
Labor	21	6.3	7
Postconviction	11	69.3	2
Prisoner Rights	23	29.0	3
Securities	12	1.7	4
Tax	3	1.1	0
Torts	22	10.9	7
Total	*156*	—	*46*

Across the sample, the mean Yearly Cases per Judgeship observed is slightly higher than the mean in the population; circuits with more cases in a given subject matter are more likely to be involved in a circuit split in that subject. Notably, there are only three tax observations, none of which come from

partially specialized circuits. In fact, the three circuits for which the data included tax observations average 1.00, 1.02, and 1.12 tax cases per judgeship—a variation unlikely to lead to different levels of expertise. The tax observations are dropped in later subject-by-subject analysis but are retained in pooled analyses.

IV. RESULTS

This Part begins with the results separated by subject matter. Simple trends in the subject-by-subject results generally support the conclusion that partial specialization improves judicial performance. When aggregated, partial specialization and judicial performance again have a positive relationship, but the relationship is not statistically significant for either independent variable. The statistical insignificance of the findings is unsurprising given that the data set includes relatively few observations. This Part concludes by quantifying the predicted effects of partial specialization; the predictions are not statistically significant, but may forecast the effects that would be observed with more data. The nontrivial magnitudes estimated suggest that partial specialization matters.

A. *Effects of Partial Specialization by Subject*

I first separate the decisions the Supreme Court agreed with from those it disagreed with and plot the average Yearly Cases per Judgeship for each set in Figure 1. For six of the seven subject matters, the circuits that the Supreme Court agreed with averaged more cases in that subject. Unless Yearly Cases per Judgeship and Supreme Court agreement rate are related, such consistency across subject matters is unlikely (but not so unlikely that the result is statistically significant).[101]

101. In Figure 1, six of the seven subject matters converge on the same relationship between Supreme Court agreement and Yearly Cases per Judgeship: the circuits that the Court agrees with average higher Yearly Cases per Judgeship. If assigned randomly with equal probability, all or all but one of the seven subject matters would converge on the same relationship 12.5% of the time. They would converge on the above relationship (higher Yearly Cases per Judgeship for the circuit that the Court agrees with) 6.25% of the time. Neither falls below the 5% standard typically required for statistical significance.

FIGURE 1
Supreme Court Agreement Versus Yearly Cases per Judgeship, by Subject Matter

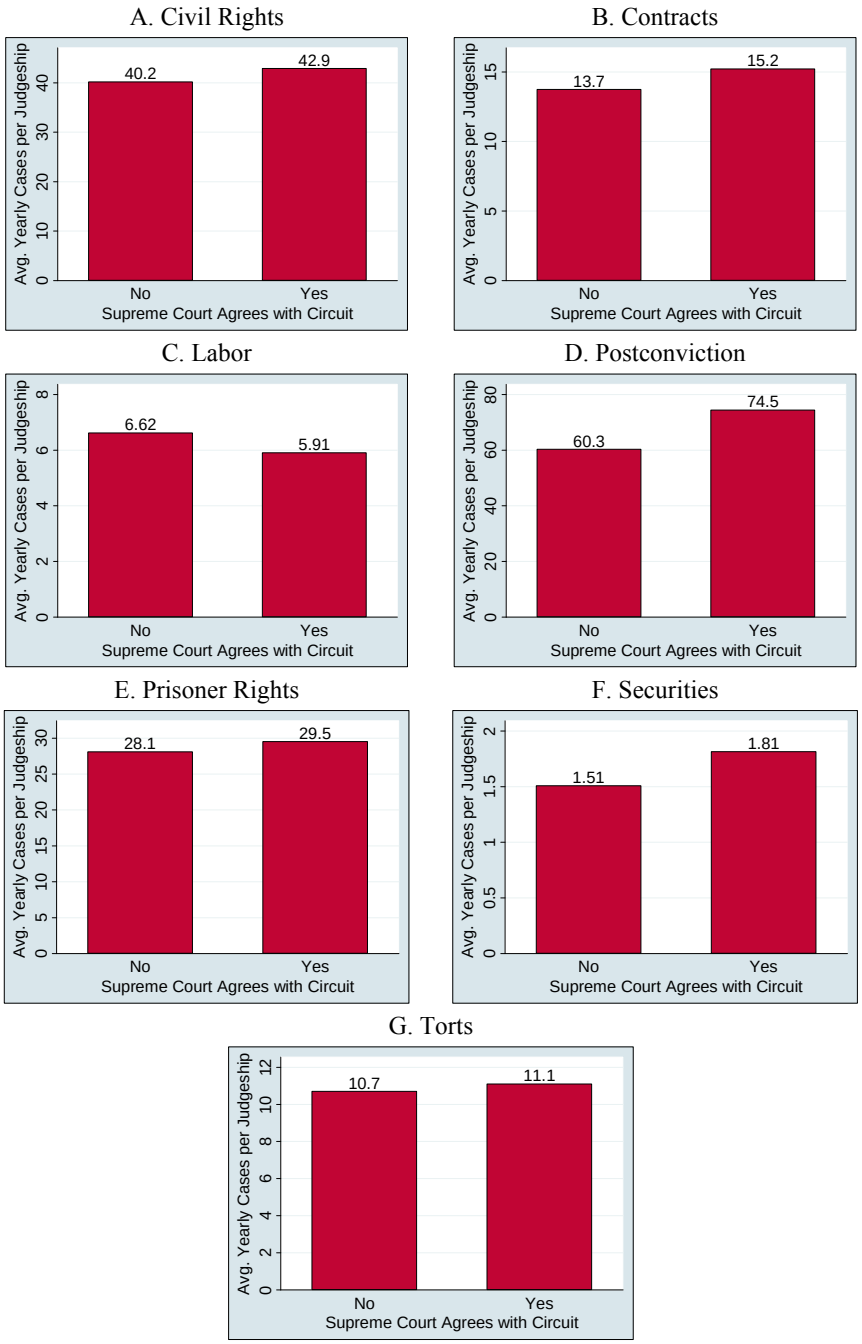
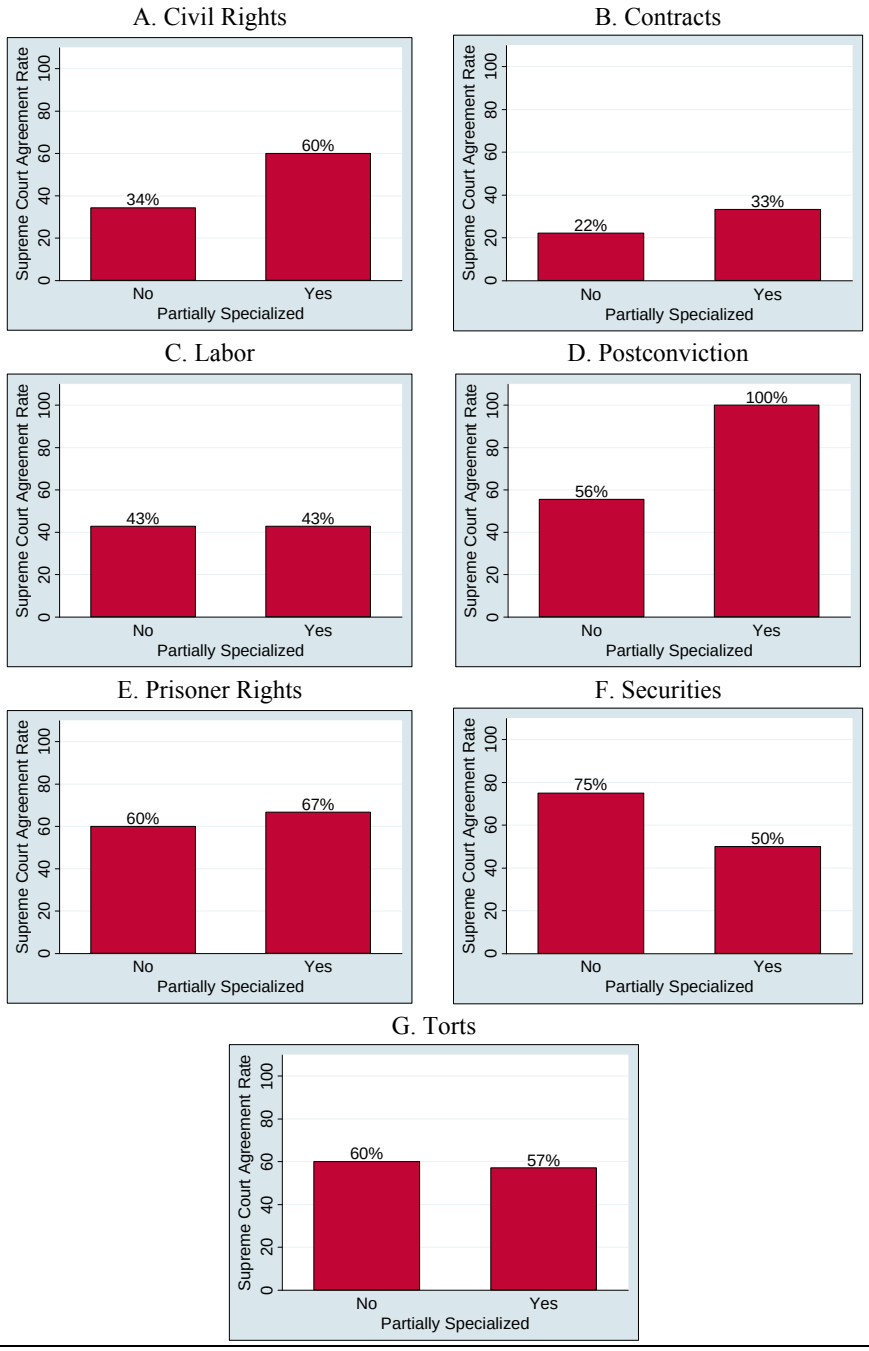


Table 4 presents Supreme Court agreement rates for each subject matter, sorted by Partially Specialized. Figure 2 plots the same information visually. Here the evidence that partial specialization improves judicial performance is thinner. The Court agrees at least as often with the nonspecialized courts in three of the seven subjects. But in one of the three (labor) the agreement rate is equal for partially specialized and nonspecialized courts, and in another (torts) the difference in agreement rate is slight. For the four subjects in which the Court agrees more often with the partially specialized court, the partially specialized courts perform far better than the nonspecialized courts do. Sorting Partially Specialized by subject matter consequently offers additional, but far fainter, support for partial specialization improving judicial performance.

TABLE 4
Supreme Court Agreement with Partially Specialized and Nonspecialized Courts, by Subject Matter

Subject Matter	Number of Decisions	Decisions the Court Agrees with	% Decisions the Court Agrees with
Civil Rights	*52*	*23*	*44%*
Partially specialized	20	12	60%
Nonspecialized	32	11	34%
Contracts	*12*	*3*	*25%*
Partially specialized	3	1	33%
Nonspecialized	9	2	22%
Labor	*21*	*9*	*43%*
Partially specialized	7	3	43%
Nonspecialized	14	6	43%
Postconviction	*11*	*7*	*64%*
Partially specialized	2	2	100%
Nonspecialized	9	5	56%
Prisoner Rights	*23*	*14*	*61%*
Partially Specialized	3	2	67%
Nonspecialized	20	12	60%
Securities	*12*	*8*	*67%*
Partially specialized	4	2	50%
Nonspecialized	8	6	75%
Tax	*3*	*1*	*33%*
Partially specialized	0	—	—
Nonspecialized	3	1	33%
Torts	*22*	*13*	*59%*
Partially specialized	7	4	57%
Nonspecialized	15	9	60%

FIGURE 2
Differences in Supreme Court Agreement Rates
from Partial Specialization, by Subject Matter



B. *Effects of Partial Specialization, Aggregated*

To test for statistical significance, I aggregated the data from the different subject matters. Aggregation is straightforward for Partially Specialized because the variable has the same scale for each subject. The relationship between Partially Specialized and Supreme Court agreement rate remains positive, but it is not statistically significant.[102] Table 5 shows this relationship.

TABLE 5
Aggregated Decisions for Partially Specialized and Nonspecialized Courts

	Supreme Court Agrees			
Circuit Type	Yes	No	Total	%
Partially Specialized	26	20	46	56.5%
Nonspecialized	52	58	110	47.3%
Total	78	78	156	50.0%

Yearly Cases per Judgeship has a different scale across subject matters (e.g., the highest Yearly Cases per Judgeship in securities would be the lowest in torts), so I could not simply pool all observations. Instead I standardize Yearly Cases per Judgeship in each subject matter to create the same scale across subjects.[103] As Table 6 shows, circuits with which the Supreme Court agrees have a higher average standardized Yearly Cases per Judgeship. Once again, a *t*-test cannot reject that the association occurs by chance.[104]

102. The null hypothesis is that the Supreme Court agreement rate will be equal for partially specialized and nonspecialized courts. A *t*-test gives $t = 1.05$, failing to reject the null hypothesis.

103. Standardizing means subtracting the population mean in a given subject from the particular circuit's Yearly Cases per Judgeship, then dividing by the population standard deviation of the subject. The mean of a standardized variable is zero and the standard deviation is one. Table S1 summarizes the observed standardized variable.

TABLE S1
Summary Statistics for Observed Standardized Yearly Cases per Judgeship

Observations	Mean	Standard Deviation	Min	Max
156	0.23	0.89	−1.65	2.64

104. The null hypothesis is that the standardized variable for circuits with which the Supreme Court agrees and with which it disagrees will be the same. A *t*-test gives $t = 0.83$, failing to reject the null hypothesis.

TABLE 6
Aggregated Decisions for Standardized Yearly Cases per Judgeship

	Observations	Mean	Standard Deviation
Supreme Court Agrees with Circuit	78	0.29	0.99
Supreme Court Disagrees with Circuit	78	0.17	0.78

C. *Quantifying the Effect of Partial Specialization*

While no single test exhibits statistical significance, each is consistent with partial specialization improving judicial performance. To offer some idea of the magnitude of this impact, this Subpart estimates the effects of increasing partial specialization in a subject. Again, none of the estimates are statistically significant and they can provide only rough guidance.

Partially Specialized cannot estimate the effects of changes in specialization because it is insufficiently fine grained. Only Yearly Cases per Judgeship helps to quantify the effects of changes in partial specialization. It answers questions at the margin: how does adding one additional case in a subject matter affect the Supreme Court agreement rate? I use a linear probability model (LPM) in these regressions.[105]

The coefficients in Table 7 predict the change in Supreme Court agreement rate for a given subject matter if a circuit adds an additional case per judgeship. The most accurate predictions fall near the mean, given in the fourth column.

105. The LPM makes fewer parametric assumptions than a probit model, and its coefficients can be interpreted more easily. *See* JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS: A MODERN APPROACH 241-45 (2d ed. 2003) (explaining LPMs); *id.* at 553-65 (explaining probit models). For extreme values the LPM can predict probabilities outside the unit interval. *See id.* at 243. But this problem will not arise for common values of the independent variable (including those near the mean). *See id.* I also use a linear functional form for Yearly Cases per Judgeship. The linear relationship models how an additional case should affect judicial performance because the judge learns and improves with each case.

TABLE 7
LPM-Estimated Effects of Yearly Cases per Judgeship on Supreme Court Agreement Rate, by Subject Matter

Subject Matter	Coefficient Estimate	Standard Error	Mean of Yearly Case/Judgeship	Effect of Standard Deviation Change
Civil Rights	0.7%	0.7%	36.4	7.7%
Contracts	1.8%	3.5%	13.5	7.2%
Labor	−5.0%	6.0%	6.0	−11.0%
Postconviction	0.6%	0.6%	67.1	16.4%
Prisoner Rights	0.4%	1.1%	30.1	4.8%
Securities	6.3%	14.2%	1.5	7.6%
Torts	1.0%	3.5%	9.9	3.2%

The coefficient estimates have straightforward interpretations. For example if a circuit increases Yearly Cases per Judgeship in contracts from thirteen to fourteen, Table 7 predicts its Supreme Court agreement rate would increase by 1.8%. To help interpret the practical meaning of the percentages, the final column of Table 7 gives the predicted effect of a one standard deviation change in Yearly Cases per Judgeship.[106] The estimates cluster around 7%. A 7% change in Supreme Court agreement rate is considerable; the difference in agreement rate between the best performing circuit and the average circuit is only 10.9%.[107] The magnitude of the (statistically insignificant) predictions suggests that the effects of partial specialization on judicial performance cannot be ignored.

CONCLUSION

The debate over whether to create more specialized courts[108] falls beyond the scope of this Note. But the finding that judicial specialization appears to have tangible expertise benefits may lend support to reforms that try to bring about specialization. At the least, the findings help to weigh the costs and benefits of specialization. A small data set limited this Note's results; future work with more data could help even more.

The added possibility of partial specialization, as opposed to the simple binary choice between specialized and nonspecialized, may add further possibilities to the debate. Perhaps the best system is the one we have: generalist judges who partially specialize in a few subject matters.

106. Table 2 lists the standard deviation for each subject.

107. Within my observations the Supreme Court agreed with the Fifth Circuit 64.9% of the time, and with the average circuit 54.0%. *See supra* Table 1.

108. *See supra* Part I.A.

Policy reforms aside, this Note's most important contribution is recognizing that the Supreme Court's resolutions of circuit splits provide a way to measure judicial performance—a tool with several advantages over reversal rates. Going forward, circuit split resolutions can enhance assessments of circuit performance. The differences between reversal rates and Supreme Court agreement rate in circuit splits, which appear in Table 1, warrant further attention. Are the differences fully attributable to the biases of reversal rates detailed in Part II.B? Or do reversal rates reflect an important piece of circuit performance that occurs outside of the context of circuit splits? Might reversal rates reflect some level of error correction by the Supreme Court? Answering these questions will illuminate how best to use reversal rates and circuit split resolutions in future research. As empirical analysis moves to the forefront of legal scholarship, data on circuit split resolutions could prove invaluable in answering questions that demand measures of judicial performance.

APPENDIX: DATA SOURCES AND COLLECTION

A. *Identifying Circuit Splits*

To find the circuit splits, I used the SCT database on Westlaw. After much trial and error, I settled on the search: <(division divide! conflict! split inconsistent differ! disagree! uncertain!) /p ("courts of appeals" circuits lower)>. The search reflects every formulation I could find that the Court uses to announce a resolution of a circuit split.[109] I did not, however, read every page of the *United States Reports*; the search probably omitted some resolutions.

I included instances when the Court identifies the existence of a circuit split and points to another case that lays out the split but does not itself name the cases.[110] In these instances the Court still acknowledges a split.

B. *Constructing Yearly Cases per Judgeship*

The numerator of Yearly Cases per Judgeship is the average number of cases each circuit had in a given subject matter across the three twelve-month periods between April 1, 2005, and March 31, 2008. I used these years because they were the most common among the circuit cases included in the splits I examined.

Ideally, I would have also included caseloads from April 1, 2004, to March 31, 2005, but the Administrative Office modified criminal offense categories in 2005 so that earlier data could not be compared with current data.[111] To check the effect of the omission, I compared the levels of circuit relative specialization in the 2005-2008 data with the levels in 2002-2003 and 2004-2005. The levels of relative specialization were stable on the whole, indicating that the additional data would have had little effect. Although some cases in the circuit splits came from earlier years, using recent data still seemed fair. A circuit can always overrule its own cases (through en banc review), and each probably considered doing so: losing litigants surely pointed to the circuit split in requesting reconsideration of the circuit's position. Declining to reconsider or overrule an earlier case is similar to adopting and accepting the new case.

---

109. *See also supra* note 36 (giving examples of the Supreme Court identifying circuit splits it is resolving).

110. *See, e.g.*, Engquist v. Or. Dep't of Agric., 553 U.S. 591, 596-97 (2008) ("The court below also acknowledged that other Circuits had applied *Olech* in the public employment context, *id.*, at 993 (citing cases), but it disagreed with those courts . . . . We granted certiorari to resolve this disagreement in the lower courts . . . ." (citing Engquist v. Or. Dep't of Agric., 478 F.3d 985 (9th Cir. 2007))).

111. *See* ADMIN. OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS: MARCH 31, 2006, at 34 (2006), *available at* http://www.uscourts.gov/uscourts/Statistics/FederalJudicialCaseloadStatistics/2006/tables/B07Mar06.pdf. In the final analysis, I ultimately omitted criminal cases. *See supra* text accompanying notes 93-94.

The denominator of Yearly Cases per Judgeship is the number of authorized judgeships for the circuit. A better denominator would have also accounted for senior judges (adjusted for each judge's caseload), visiting judges (typically from district courts), and vacancies. But each varies through the time period, and gauging visiting judge participation or senior judge caseload would be challenging.

As an alternative denominator, I divided the numerator by population residing in the circuit.[112] Circuit population has no direct role in measuring specialization and has no place outside a test for robustness. But it may indicate when the number of authorized judgeships inadequately reflects caseload in the circuit; a circuit may rely more on senior and visiting judges because it has too few authorized judgeships. I reran the tests in Part IV.A with population as the denominator, and no results changed.

Most of the data used to construct Yearly Cases per Judgeship came from the Administrative Office of the Courts. All civil cases filed in a federal court are coded with a Nature of Suit (such as "Antitrust" or "Fair Labor Standards Act"). I aggregated the individual Nature of Suit categories into subject matters following aggregations by the Administrative Office[113] and other scholars.[114] The Administrative Office annually provides the number of cases in each Nature of Suit or Offense for each regional court of appeals.[115]

112. I used the 2000 census populations. *See Your Gateway to Census 2000*, U.S. CENSUS BUREAU, http://www.census.gov/main/www/cen2000.html (last visited Jan. 5, 2011).

113. *See* ADMIN. OFFICE OF THE U.S. COURTS, CIVIL COVER SHEET (2007), *available at* http://www.uscourts.gov/uscourts/FormsAndFees/Forms/JS044.pdf.

114. *See* Cheng, *supra* note 5, app. B.

115. *See* ADMIN. OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS: MARCH 31, 2008, at 31 tbl.B-7 (2008), *available at* http://www.uscourts.gov/uscourts/Statistics/FederalJudicialCaseloadStatistics/2008/tables/B07Mar08.pdf; ADMIN. OFFICE OF THE U.S. COURTS, FEDERAL JUDICIAL CASELOAD STATISTICS: MARCH 31, 2007, at 31 tbl.B-7 (2007), *available at* http://www.uscourts.gov/uscourts/Statistics/FederalJudicialCaseloadStatistics/2007/tables/B07Mar07.pdf; ADMIN. OFFICE OF THE U.S. COURTS, *supra* note 111, at 30 tbl.B-7.