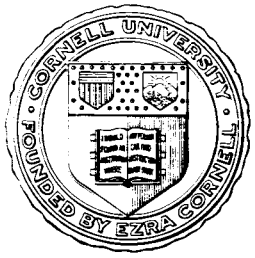
Cornell
Journal of Law
and Public Policy




Volume 10 Spring 2001 Number 2

# STRANGE BEDFELLOWS: POLITICS, COURTS, AND STATISTICS: STATISTICAL EXPERT TESTIMONY IN VOTING RIGHTS CASES

*WENDY K. TAM CHO*
*ALBERT H. YOON*

# STRANGE BEDFELLOWS: POLITICS, COURTS, AND STATISTICS: STATISTICAL EXPERT TESTIMONY IN VOTING RIGHTS CASES

*by Wendy K. Tam Cho* and Albert H. Yoon***



Issues surrounding redistricting and the Voting Rights Act of 1965[1] raise interesting methodological, political, and legal questions. The political ramifications are obvious, since redistricting affects how residents of different ethnic and racial groups are organized within voting districts. In order to assess whether the existing district lines must be redrawn to comply with civil rights legislation, courts must often analyze some set of population data with advanced statistical techniques. The problem, however, is that statistics is not a perfect science. Since the statistical methods employed to evaluate the fairness of a redistricting scheme are subject to factors of uncertainty, assessing statistical analysis properly is sometimes an art unto itself.

---

* Department of Political Science and Department of Statistics, University of Illinois at Urbana-Champaign, 361 Lincoln Hall, 702 S. Wright St., Urbana, IL 61801-3696.

** Northwestern University School of Law, Department of Political Science (by courtesy), Northwestern University.

We would like to thank Bruce Cain, Brian Gaines, and Morgan Kousser for helpful comments. Any errors are the responsibility of the authors.

[1] Pub. L. No. 89-110, 79 Stat. 437 (1965) (codified as amended at 42 U.S.C. §§ 1973-1973bb (1994)).

Because of these inherent complexities, expert witnesses may be called upon to help bridge the chasm between data analysis and the resulting court decision or subsequent rule of law. Expert testimony is often critical as these cases frequently turn on the credibility of the expert witnesses and the clarity of their testimony. However, the degree to which judges competently review this testimony varies considerably. Obviously, when judicial competence is called into question, so too is the legal process.

Decisions, then, which lack a competent evaluation of expert testimony, must necessarily turn on less idealistic and more ideological grounds. The questionability of high standards rises to the forefront as the tension between science and law increases and becomes more visible. As Judge Learned Hand once wrote,

> The trouble with [expert testimony] is that it is setting the jury to decide, where doctors disagree . . . . But how can the jury judge between two statements founded on an experience admittedly foreign in kind to their own? It is just because they are incompetent for such a task that the expert is necessary at all.[2]

As science progresses, and as new statistical methods are devised, Judge Hand's comments become increasingly pertinent. Judges grow less and less competent to judge technical information with every new statistical advance. The resulting case law becomes less objectively based on reality and reason and more tied to a judge's ability to justify subjective ideological stances or to the persuasive skills of attorneys and expert witnesses. The rapid progress in the mathematical sophistication of data analysis, then, signals a need to revisit the manner in which judges decide cases.

There are other potentially fundamental flaws in the legal process besides the judging of expert testimony. For example, a competent attorney can always blur the line between statistical evidence and law. Alternatively, the law can be faulty because it is easily distorted by wealth. Few would dispute that the quality of legal representation afforded or provided for the poor is markedly inferior to that of the wealthy.[3] The United States Constitution attempts to mitigate the effects of wealth in matters of criminal law by guaranteeing an indigent defendant's right to

---

[2] Learned Hand, *Historical and Practical Considerations Regarding Expert Testimony*, 15 HARV. L. REV. 40, 54 (1901).

[3] *See Conflicts of Interest in the Legal Profession*, 94 HARV. L. REV. 1244, 1401–02 (1981) (stating that "most compelling is the overwhelming lack of adequate legal representation for America's poor").

effective assistance of counsel.[4] This standard, however, remains relatively low.[5] More importantly, it does not extend to civil matters. While a civil litigant provided with ineffective counsel can pursue a subsequent separate tort action against his attorney, it will not necessarily remedy the harm done to the case at hand. These issues are outside the scope of this article, but are addressed by other legal scholars.[6] Here, we will concentrate strictly on the interplay between expert witnesses and judges. We view this as the critical interaction and one that can most easily be altered to fit the ideals of the legal process.

Our analysis focuses specifically on statistical evidence in Voting Rights cases. In this arena there are two separate but intertwined sets of issues to explore. The first involves the legal question of whether the judge and the courtroom are the proper venues for these types of decisions. The second concerns the complexity that is often inherent in data analysis. We first discuss the precedents for reviewing scientific evidence in court cases. In doing so, we highlight specific problems with the current framework for admitting statistical expert testimony. Next, we examine the data and the methods suggested by the different expert witnesses in past court cases. We then discuss the reasoning and decisions of the judges. It will become evident that the decisions in these cases often rely upon plainly faulty reasoning and misunderstanding rather than on sound scientific inquiry. Lastly, we present recommendations for producing a system that is better suited to sort through the complexity inherent in these cases.

## I. THE ISSUES OF LAW

The general principle underlying scientific evidence has remained unchanged throughout this century. For scientific evidence to be admissible, as with all evidence, it must be deemed both reliable and relevant. Legal scholars, such as Learned Hand, viewed relevance and reliability

---

[4] U.S. CONST. amend. VI (stating that "In all criminal prosecutions, the accused shall enjoy the right to . . . the Assistance of Counsel for his defense."); *see also* U.S. CONST. amend. XIV (extending the Sixth Amendment to the states). The Supreme Court has extended the Sixth Amendment to include the effective assistance of counsel to proceedings prior to sentencing, *see* Gardner v. Florida 430 U.S. 349, 357 (1977); Mempa v. Rhay, 389 U.S. 128 (1967); Specht v. Patterson, 386 U.S. 605 (1967); and the first appeal following judgment, *see* Douglas v. California, 372 U.S. 353 (1963).

[5] *See, e.g.*, Stephen B. Bright, *Casualties of the War on Crime: Fairness, Reliability, and the Credibility of Criminal Justice Systems*, 51 U. MIAMI L. REV. 413 (1997) (discussing, in part, how criminal defendants are subjected to ineffective assistance of counsel, both at trial and on appeal).

[6] *See, e.g.*, Meredith J. Duncan, *Legal Malpractice by Any Other Name: Why a Breach of Fiduciary Duty Claim Does Not Smell as Sweet*, 34 WAKE FOREST L. REV. 1137 (1999); Manuel R. Ramos, *Legal Malpractice: No Lawyer or Client is Safe*, 47 FLA. L. REV. 1 (1995) (looking at the issue from an empirical perspective).

not as two distinct concepts, but rather the former being a subset of the latter.[7] Scientific evidence in the courtroom has changed, however, because the courts have reinterpreted the reliability standard of scientific evidence on several occasions, and because the courts have reconsidered the optimal actor for determining whether the evidence is reliable.

For the greater half of this century, courts adopted a "general admissibility" standard for scientific evidence, pursuant to the District of Columbia Court of Appeals decision in *Frye v. United States*.[8] In *Frye*, the defendant was convicted of second-degree murder. During his trial, the defendant attempted to introduce evidence of a test based on his systolic blood pressure.[9] He argued that the test, a primitive form of lie detector, would exonerate him by distinguishing changes in his blood pressure reflecting general anxiety from changes reflecting a false statement.[10] The trial court refused to admit the test and the Court of Appeals affirmed, reasoning,

> Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained *general acceptance* in the particular field in which it belongs.
>
> We think the systolic blood pressure deception test has not yet gained such standing and scientific recognition among physiological and psychological authorities as would justify the courts in admitting expert testimony deduced from the discovery, development, and experiments thus far made.[11]

Accordingly, *Frye* required that scientific evidence must first be rooted in a theory that was developed, reviewed, and validated by the relevant academic community. In the aftermath of *Frye*, the "general acceptance" test became the dominant standard for determining the admissibility of novel scientific evidence at trial.[12] Though generally embraced, *Frye*

---

[7] *See* Hand, *supra* note 2, at 52.

[8] 293 F. 1013 (D.C. Cir. 1923).

[9] *Id.* at 1014.

[10] *Id.* at 1013.

[11] *Id.* at 1014 (emphasis added).

[12] *See* Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 586 (1993) (citing E. GREEN & C. NESSON, PROBLEMS, CASES, AND MATERIALS ON EVIDENCE 649 (1983)); *see also* Edward

drew criticism by some scholars who viewed the "general acceptance" standard as imposing too high a standard, excluding otherwise helpful information for the jury. According to Judge Harvey Brown,

> [T]he *Frye* test was criticized because the newness of a scientific theory does not necessarily reflect its unreliability, "nose counting" of the scientific community could be difficult and unhelpful, and the standard delays the admissibility of new evidence simply because the scientific community has not had adequate time to accept the new theory.[13]

In 1975, Congress codified the Federal Rules of Evidence,[14] which provided a uniform set of evidence rules for civil and criminal cases tried in federal court. In response to the criticism of the *Frye* standard, Congress included Rule 702 to address the admissibility of expert testimony.[15] The specific text reads, "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise."[16] Although neither the rule nor the commentary notes refer to *Frye*, the rule emphasizes relevance rather than reliability or "general acceptance" of the scientific evidence.[17] Nonetheless, many federal courts continued to follow *Frye* in evaluating scientific evidence.[18]

At first, the Supreme Court appeared to embrace the idea that Rule 702 focused on the relevance of scientific testimony and left the issue of reliability to the jury. In *Barefoot v. Estelle*,[19] the Court considered a habeas corpus petition of a defendant convicted of capital murder in Texas. During sentencing to determine whether the defendant was eligi-

---

J. Imwinkelried, *The Standard for Admitting Scientific Evidence: A Critique from the Perspective of Juror Psychology*, 28 VILL. L. REV. 554, 557 (1983) (stating that by the mid-1970s, approximately forty-five states followed the *Frye* standard for admissibility).

13 Harvey Brown, *Eight Gates for Expert Witnesses*, 36 HOUS. L. REV. 743, 779 (1999); *See also* CHARLES T. MCCORMICK, EVIDENCE § 203, at 491 ((2d ed. 1972); P. GIANELLI & E. IMWINKELREID, SCIENTIFIC EVIDENCE § 1–5(G), at 27–30 (3d ed. 1999).

14 *See* Pub. L. No. 93–595, 88 Stat. 1926, 1939 (1975) (codified as amended at 28 U.S.C. app. 775–ff (1999)).

15 *See* FED. R. EVID. 702.

16 FED. R. EVID. 702.

17 *See Daubert*, 509 U.S. at 589 (stating that "under the [Federal Rules of Evidence] the trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable.").

18 *See* Gordon J. Beggs, *Novel Expert Evidence in Federal Civil Rights Litigation*, 45 AM. U. L. REV. 1, 20 (1995) (citing cases illustrating circuit courts' reluctance to overturn *Frye*, and instead "incorporat[ing] the general acceptance standard into the relevance determination under Rule 702.").

19 463 U.S. 880 (1983).

ble for the death penalty, the jury was asked to consider the probability that he might commit another violent act.[20] Toward that end, the trial court allowed psychiatric testimony by the state's experts averring that such a probability existed. The Court held that the trial court did not violate the criminal defendant's due process rights by admitting psychiatric testimony regarding his future dangerousness. It stated,

> We are unconvinced . . . that the adversary process cannot be trusted to sort out the reliable from the unreliable evidence and opinion about future dangerousness, particularly when the convicted felon has the opportunity to present his own side of the case . . . . Petitioner's entire argument . . . is founded on the premise that the jury will not be able to separate the wheat from the chaff. We do not share in this low evaluation of the adversary process.[21]

Moreover, the Court, guided by existing rules of evidence including Rule 702, stated that "the rules of evidence generally extant at the federal and state levels anticipate that relevant, unprivileged evidence should be admitted and its weight left to the fact finder, who would have the benefit of cross-examination and contrary evidence by the opposing party."[22] Despite its bold position on the admissibility of expert testimony, *Barefoot* was subsequently cited more for the proposition that a habeas petition need show only that the issues raised are "debatable among jurists of reason."[23]

Ten years later, the Supreme Court shifted course and expressed concern with the reliability of scientific evidence. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[24] involved a case where the plaintiffs—a mother and her two children—alleged that the children's serious birth defects were caused by the mothers' prenatal ingestion of Bendectin, a prescription drug marketed by the defendant. In making their argument, the plaintiffs sought to include expert testimony of eight experts who testified that Bendectin caused birth defects in animals.[25] The defendant countered with its own expert who testified that no study showed that the drug caused birth defects in humans.[26] The district court applied the

---

[20] *Id.* at 884.

[21] *Id.* at 901.

[22] *Id.* at 898; *see also* Michael H. Gottesman, *From Barefoot to Daubert to Joiner: Triple Play or Double Error?*, 40 ARIZ. L. REV. 753, 755 (1998) (discussing the *Barefoot* decision).

[23] 463 U.S. at 893 n.4.

[24] 509 U.S. 579 (1993).

[25] *Id.* at 583.

[26] *Id.* at 582.

*Frye* standard,[27] holding that the plaintiffs' experts' testimony did not satisfy the "general acceptance" standard, and the Ninth Circuit agreed.[28]

The Supreme Court, however, reversed and remanded to the Ninth Circuit. The Court expressly stated that *Frye* was superceded by Rule 702, which made no mention of the "general acceptance" standard.[29] At first impression, it appeared that the Court was merely eliminating any residual loyalty that lower courts may have retained for the *Frye* standard. At the same time, however, the Court also held that the Federal Rules of Evidence compel a trial judge to "ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable."[30] The Court considered four factors—not meant to be exhaustive—which courts could consider to determine the reliability of a particular scientific theory or technique: testability, peer review, error rates, and "acceptability" in the relevant scientific community.[31]

Ultimately, *Daubert* was as much about backing away from *Barefoot* as it was about removing courts' loyalty toward *Frye*. The faith in juries that the Court articulated in *Barefoot* was now replaced by a skepticism in the competency of juries to process scientific information. When considering the threat that judges would exclude valid scientific testimony, the Court employed a cost-benefit argument, and reasoned that judges were in a better position than juries to determine the reliability of the expert's testimony.

> We recognize that, in practice, a gatekeeping role for the judge, no matter how flexible, inevitably on occasion will prevent the jury from learning of authentic insights and innovations. That, nevertheless, is the balance that is struck by Rules of Evidence designed not for the exhaustive search for cosmic understanding but for the particularized resolution of legal disputes.[32]

This assertion contrasted boldly with *Barefoot*, and was beyond the statutory language of Rule 702. Chief Justice Rehnquist, in his partial dissent, questioned the majority's foundation for such a belief, stating that "scientific knowledge, scientific method, scientific validity, and peer review . . . [are] matters far afield from the expertise of judges."[33] Interestingly, neither the majority nor the concurrence/dissent made any mention of *Barefoot*.

---

27 *See* Daubert v. Merrell Dow Pharm., Inc., 727 F. Supp. 570 (S.D. Cal. 1989).

28 *See* Daubert v. Merrell Dow Pharm., Inc., 951 F.2d 1128 (9th Cir. 1991).

29 509 U.S. at 587.

30 *Id.* at 589.

31 *Id.* at 593–94.

32 *Id.* at 597.

33 *Id.* at 599 (Rehnquist, C.J., concurring in part, dissenting in part).

In 1997, the Court in *General Electric Co. v. Joiner*[34] further empowered the district court to determine the admissibility of scientific evidence. The case involved a plaintiff suffering from small-cell lung cancer who alleged that his disease was promoted by workplace polychlorinated biphenyls (PCBs) and other toxins.[35] In making his case, the plaintiff sought to present experts who testified that PCBs and the other toxins in question could cause lung cancer, and were likely the cause of the plaintiff's cancer.[36] The district court, in granting summary judgment for the defendant, disallowed the testimony, claiming that the plaintiff's experts' testimony did not rise above "subjective belief or unsupported speculation."[37] The Eleventh Circuit reversed the district court, reasoning that "[b]ecause the Federal Rules of Evidence governing expert testimony display a preference for admissibility, we apply a particularly stringent standard of review to the trial judge's exclusion of expert testimony."[38] The Supreme Court rejected the Eleventh Circuit's heightened scrutiny, whose reliance on *Daubert* for its standard of review appeared wholly unsupported.

> On a motion for summary judgment, disputed issues of fact are resolved against the moving party—here, petitioners. But the question of admissibility of expert testimony is not such an issue of fact, and is reviewable under the *abuse of discretion standard.*
>
> We hold that the Court of Appeals erred in its review of the exclusion of Joiner's experts' testimony. In applying an overly "stringent" review to that ruling, it failed to give the trial court the deference that is the hallmark of abuse-of-discretion review.[39]

To eliminate any residual doubt, the Court noted that pursuant to an abuse of discretion standard, "the appellate court will not reverse in such a case, unless the ruling is manifestly erroneous."[40]

Thus, since *Frye*, the Court has vacillated with respect to its approach to scientific evidence. *Frye* required "general acceptability," which served as the standard for much of this century. The codification of Rule 702 in 1975 shifted the focus of scientific evidence back to relevance, and dealt with reliability only implicitly. The Supreme Court ini-

---

34 522 U.S. 136 (1997).

35 *Id.* at 139.

36 *Id.*

37 Joiner v. Gen. Elec. Co., 864 F. Supp. 1310, 1326 (N.D. Ga. 1994).

38 Joiner v. Gen. Elec. Co., 78 F.3d 524, 529 (11th Cir. 1996) (citing *Daubert*, 509 U.S. at 588; *In re* Paoli R.R. Yard PCB Litig., 35 F.3d 717, 750 (3d Cir. 1994)).

39 522 U.S. at 143 (emphasis added).

40 *Id.* at 142 (quoting Spring Co. v. Edgar, 99 U.S. 645, 658 (1879)).

tially adopted a plain reading for Rule 702 in *Barefoot*, leaving the issue of reliability to the jury. In *Daubert*, however, the Court backed away from *Barefoot*, and dealt with the issue of reliability explicitly by vesting that determination in the trial judge. *Daubert* was also a bit disingenuous—while explicitly holding that Rule 702 superceded *Frye's* "general acceptability" standard, it reintroduced the factor of acceptability, albeit a less exacting standard. *Joiner* simply reinforced the authority of the trial judge in matters of expert testimony by providing a deferential abuse of discretion standard. Most recently, the Court in *Kumho Tire Company, Ltd. v. Carmichael* simply extended this discretion to all expert testimony.[41]

## II. THE CHALLENGES JUDGES FACE WHEN DEALING WITH SCIENTIFIC INFORMATION

In general, judges are well-suited to address the legal issues that arise with ensuring fair political representation, which is at the heart of the Voting Rights Act of 1965; however they are ill-suited to render decisions that require statistical sophistication. Unfortunately, this sophistication is often necessary to establish the facts of the case. As Justice Frankfurter noted,

> So many cases turn on the facts, principles of law not being in controversy. It is only when the facts are interwoven with the questions of law which we should review that the evidence must be examined and then only to the extent that it is necessary to decide the questions of law.[42]

The inability to separate the facts from the ruling of law is particularly true in Voting Rights cases.

### A. THE ATTRIBUTES AND LIMITATIONS OF JUDGES

In the legal profession, federal judges are highly regarded. The majority have achieved prominence in the lower courts, one of the central factors in their nomination to the federal bench. In President Clinton's first term in office, for example, roughly half of the confirmed nominees for federal district court were already judges at the state level, and over a third of them had spent time as prosecutors, many of them as a district attorney or as a United States attorney;[43] the percentages of former

---

41 526 U.S. 137 (1999).

42 Dick v. N.Y. Life Ins. Co., 359 U.S. 437, 455 (1959) (J. Frankfurter, dissenting). On the interplay between facts and law, Robert Weisberg of Stanford Law School has commented that "cases turn on highly fact-specific rules of law."

43 Sheldon Goldman & Elliot Slotnick, Clinton's First Term Judiciary: Many Bridges to Cross, 80 JUDICATURE 254, 258 (1997).

judges and prosecutors confirmed to the court of appeals was even higher.[44] A considerable percentage, particularly in the court of appeals, attended the nation's top law schools.[45] The confirmation process, as well, helps to ensure that only qualified candidates elevate to the bench. On the federal level, all Article III judges must be nominated by the President and confirmed by the Senate.[46] Although a judicial appointment requires sponsorship by the Senator or local party leaders and judicial appointments are often perceived as partisan,[47] political connections cannot trump legal competence.[48] Members of the minority party as well as the media scrutinize nominees, and the American Bar Association (ABA) evaluates each judicial nominee and gives them a rating of "well qualified," "qualified," and "unqualified."[49] Most nominees receive a rating of "Well Qualified."[50]

Despite her stature in the legal community, the average judge specializes in one area of law, and is not an expert in all areas of the law. A judge with an expertise in, say, criminal law is unlikely to possess the same breadth of knowledge in the areas such as tax, labor, race or gender discrimination, or civil rights. However, since judges are required to

---

[44] *Id.* at 261.

[45] For example, on the D.C. Circuit Court of Appeals, each of the twelve active circuit judges attended law schools currently ranked in the top 25 of *U.S. News and World Report's* 1999 rankings, and nine of the twelve attended laws schools ranked in the top 10. Similarly, on the Sixth Circuit Court of Appeals, ten of the eleven attended law schools ranked in the top 25, while nine of those ten attended law schools currently ranked in the top 10. See 6 JUDICIAL YELLOW BOOK (Leadership Directories, Inc. ed., 2001); *Exclusive Rankings*, U.S. NEWS AND WORLD REPORT, Mar. 29, 1999, at 90-95.

[46] *See* U.S. CONST. art. II, § 2, cl.2, which states that the President "shall nominate, and by and with the Advice and Consent of the Senate . . . Judges of the supreme Court, and all other Officers of the United States, whose Appointments are not herein otherwise provided for, and which shall be established by Law."

[47] *See, e.g.*, SHELDON GOLDMAN, PICKING FEDERAL JUDGES: LOWER COURT SELECTION FROM ROOSEVELT THROUGH REAGAN 17–30 (1997) (describing the motivation behind President Franklin D. Roosevelt's nominations to the bench).

[48] *See, e.g.*, JOHN P. FRANK, CLEMENT HAYNSWORTH, THE SENATE, AND THE SUPREME COURT 112 (1991). During the confirmation hearings for the nomination of Judge Harrold Carswell, who was held in low regard as a judge, for the Supreme Court, Carswell supporter Senator Hruska remarked, "[T]here are a lot of mediocre judges and people and lawyers. They are entitled to a little representation, aren't they, and a little chance? We can't have all Brandeises and Frankfurters and Cardozos and stuff like that there." (internal quotation marks omitted).

[49] The rating "Exceptionally Well Qualified" was dropped at the beginning of the George H. W. Bush Administration. For a succinct discussion of the federal confirmation process, *see* Goldman, *supra* note 47, at 9–14.

[50] During President Clinton's first administration, 64% of Clinton's district court nominees received a "Well Qualified" rating, while 83% of Clinton's circuit court nominees received that same rating. Of course, rational choice scholars and laypersons alike would agree that the President is a sophisticated political actor, and would generally nominate only those candidates that he believed would pass any competency requirement. *See, e.g.*, Bryon J. Moraski & Charles R. Shipan, *The Politics of Supreme Court Nominations: A Theory of Institutional Constraints and Choices*, 43 AMER. J. POL. SCI. 1069–95 (1999).

hear and decide cases involving many areas of law, over time, they become more familiar with other areas of law. Nonetheless, as Chief Justice Rehnquist pointed out in his partial dissent in *Daubert*, the legal qualifications of the average judge do not necessarily extend to other disciplines, especially in math or the hard sciences.[51] Rehnquist's claim seems intuitive since the preparation, study, and practice of law, while not hostile to these disciplines, do not require any background or training in them.[52] While the absence of formal schooling does not preclude a judge from acquiring particular skills, certain disciplines, statistics being one of them, do not lend themselves well to non-formal training. This is especially true if one does not have the requisite background or training in mathematical fundamentals.

The on-the-job training of judges, while helpful, is also limited. Once appointed, judges do attend "judge schools," sponsored by the Judicial Education Division of the Federal Judicial Center. The training session, often referred to as "Baby Judge School," is offered annually in Washington D.C. to newly appointed federal judges from across the country.[53] These sessions, however, are typically law-centered, offering brush-up courses in areas such as jurisdiction, pre-trial and trial strategy, case management, habeas corpus procedures, sentencing guidelines, pro se litigants, discrimination suits, and press relations.[54] For the 1998–99 judicial term, the Federal Judicial Center did not offer a single seminar on the use of statistics in the courtroom.[55] Alternatively, many judges take advantage of other seminars offered by organizations such as law

---

[51] Challenging the majority, Chief Justice Rehnquist wrote

> I defer to no one in my confidence in federal judges; but I am at a loss to know what is meant when it is said that the scientific status of a theory depends on its "falsifiability," and I suspect some of them will be, too. I do not doubt that Rule 702 confides to the judge some gatekeeping responsibility in deciding questions of the admissibility of proffered expert testimony. But I do not think it imposes on them either the obligation or the authority to become amateur scientists in order to perform that role. I think the Court would be far better advised in this case to decide only the questions presented, and to leave the further development of this important area of the law to future cases.

509 U.S. at 599 (Rehnquist, C.J., concurring in part, dissenting in part).

[52] The District of Columbia court, typically regarded as the elite federal court on both the district and circuit level, illustrates this point. Eleven of the twelve active district court judges do not possess an undergraduate or graduate degree in the sciences. On the circuit level, only two of the twelve active judges possess a bachelor of science degree, and none completed graduate work in the math or sciences. *See* Westlaw Directory – Judges, *at* http://www.westlaw.com.

[53] Rex Bossert, *A Week at Boot Camp for Judges*, NAT'L. L.J., July 7, 1997, at A1.

[54] *See id.*; *see also* Bruce Brown, *Where 'Baby Judges' Learn Their ABCs; Weeklong Session Introduces New Appointees to Federal Judiciary*, FULTON COUNTY DAILY REP., June 18, 1996.

[55] *See Services for Judges from the Federal Judicial Center*, 1998–1999, *at* http://www.fjc.gov.

schools. For example, Yale Law School hosts civil liability conferences aimed at the judiciary, as do Columbia and New York University law schools.[56] As with the seminars offered by the Federal Judicial Center, these courses also focus on traditional legal topics.

In 1994, motivated by *Daubert*, the Federal Judicial Center published the Reference Manual on Scientific Evidence.[57] This manual offers reference guides on common topics where law and science intersect, such as epidemiology, toxicology, survey research, forensic DNA evidence, and economic valuation in damage awards. It also includes a section on basic statistics and another on multiple regression.[58] It is unclear, however, whether a judge, without any prior formal training in statistics, could fully comprehend the material presented in these sections. Moreover, because the chapters contain only introductory material, they do not cover many of the more advanced techniques and principles presented by statisticians in trial. Lastly, these chapters, like all the chapters in the manual, serve only as a reference.[59] Judges may elect not to make use of them.

### B. THE ATTRIBUTES AND LIMITATIONS OF THE JUDGE'S CHAMBERS[60]

From an institutional perspective, it is not necessary for a judge to possess a full understanding of scientific evidence if her staff is capable of understanding the scientific evidence.[61] Her staff can analyze the testimony and inform her of robust, appropriate, and ultimately admissible methods. Justice Breyer, in his concurrence in *Joiner*, suggested that judges hire law clerks who possessed such specialized training.[62] The chambers of most judges, unfortunately, do not have such resources available. The typical chambers for a district judge consists of two clerks and a secretary, while the circuit court consists of either four clerks and a secretary or three clerks and two secretaries. The overwhelming majority of clerks are annual clerks who have recently gradu-

---

56 *See* Hon. Jack B. Weinstein, *Learning, Speaking, and Acting: What are the Limits for Judges?*, 77 JUDICATURE 322, 323 (May–June 1994).

57 Federal Judicial Center, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE (1994).

58 *Id.* David Kaye and David Freedman wrote the chapter on statistics, the latter of whom is a statistician at the University of California at Berkeley. Daniel Rubinfeld, an econometrician and law professor, also at University of California at Berkeley, wrote the multiple regression chapter.

59 *See id.* at 3.

60 Much of the information in this section is drawn from the co-author's experience while clerking for the Hon. R. Guy Cole, Jr. of the U.S. Court of Appeals for the Sixth Circuit for the 1998–99 term. The views expressed herein do not reflect the opinion of any judge on the circuit, nor do they reflect the co-author's direct experience in chambers.

61 For the sake of consistency, the pronoun reference to a judge takes on the female gender.

62 522 U.S. 136, 149–50 (1997) (Breyer, J., concurring).

ated from law school. While the typical law clerk has distinguished himself or herself in law school—thus earning a federal clerkship—anecdotal evidence suggests that he or she is no more likely than the judge to possess a technical background.[63]

### C. THE CHALLENGES FACING A JUDGE IN PREPARATION OF A CASE

In preparing for any given case, a judge must overcome an informational deficit. In the beginning, she knows much less about the facts—and in some cases, the law—than do the litigants. When the facts are not in dispute between the parties, the informational deficit can be fairly easily overcome, since the case depends solely on the application of law. Even if the facts are in dispute, however, assuming that the information sought is within the skill set of the judge or chambers, the institutional structure of the judiciary allows a judge to adjust to the increased requirements to process such information. For instance, the executive office of the court can redirect cases to other judges in the district or circuit—e.g., as was recently done for U.S. District Judge Thomas Penfield Jackson in the Justice Department's antitrust suit against Microsoft Corporation.[64] Provided that every case does not present such a strain on the judge and his staff, the institutional design of the chamber can respond adequately to labor-intensive cases.

In some cases, comprehension of the facts depends primarily on specific sets of skills that may not be possessed by either the judge or her staff. The context of Voting Rights is a case in point.[65] The Voting Rights Act prohibits the practice that denies or abridges, on grounds of race or color, the right of any citizen to vote.[66] Specifically, the Act provides plaintiffs with an equal protection claim whenever the State has

---

63 Public information on judicial law clerks is scarce. The Administrative Office of the U.S. Courts does not maintain records on annual clerks; most information about clerks are retained individually by the judges themselves. The Hon. Richard Posner of the U.S. Court of Appeals for the Seventh Circuit, however, in THE FEDERAL COURTS: CHALLENGE AND REFORM 139–59 (1996), includes a section entitled, "The Rise of the Law Clerk," in which he describes how the burgeoning case load requirements on judges has compelled them to rely more on their clerks, resulting in "colorless and plethoric" decisions from the bench. *Id.* at 146.

64 United States v. Microsoft Corp., 65 F. Supp.2d 1 (D.D.C. 1999) (issuing findings, after several months of trial, that Microsoft exercised monopoly power over its competitors in the operating systems market).

65 A lengthy discussion of Voting Rights law is beyond the scope of this article, but is addressed by several legal scholars. *See, e.g.*, Bernard Grofman & Lisa Handley, *1990s Issues in Voting Rights*, 65 MISS. L.J. 205 (1995); Samuel Issacharoff & Pamela S. Karlan, *Standing and Misunderstanding Voting Rights Law*, 111 HARV. L. REV. 2276 (1998); Pamela S. Karlan, *Defining Democracy for the Next Century Loss and Redemption: Voting Rights at the Turn of a Century*, 50 VAND. L. REV. 291 (1997); *The Supreme Court 1995 Term-Leading Cases*, 110 HARV. L. REV. 135, 185–96 (1996); *The Supreme Court 1994 Term-Leading Cases*, 109 HARV. L. REV. 111, 160–70 (1995).

66 42 U.S.C. §§ 1973-1973bb (1994).

used race as a basis for separating voters into districts.[67] After the Supreme Court held that the Voting Rights Act prohibited only intentional discrimination,[68] Congress amended Section 2 in 1982 to provide minority groups with a remedy for vote dilution without requiring them to show that the majority engaged in intentional discrimination.[69] Congress provided a non-exhaustive list of factors for courts to consider when making this determination, directing courts to engage in a "searching practical evaluation of the past and present reality" of the particular political system.[70]

Voting Rights law itself is complex, and has been evolving for the past two decades. When evaluating a vote dilution claim, courts must first look for three preconditions required by the minority: geographical compactness of the minority group; minority political cohesion; and majority bloc voting.[71] If these conditions are met, courts then must consider the totality of the circumstances.[72] The court, if it finds that race is the predominant motive in creating districts, applies strict scrutiny, and the districting plan must be narrowly tailored to serve a compelling governmental interest in order to survive review.[73] The Supreme Court has assumed, without deciding, that compliance with Section 2 of the Voting Rights Act can be a compelling state interest.[74] The laws set forth in Voting Rights cases are further complicated because typically an expert must analyze aggregated precinct data to determine whether the plaintiffs satisfy any of the preconditions.

In court, an expert testifies that a Section 2 violation occurred based on his model of the aggregated precinct data.[75] The judge likely understands the conclusions of the testimony, i.e., the data reflect that minority voters within a district are politically cohesive, or that the majority voters, voting as a bloc, defeated the minority's preferred candidate.[76] She

---

67 *See* Shaw v. Reno, 509 U.S. 630, 640, 645 (1993).

68 City of Mobile v. Bolden, 446 U.S. 55, 61–65 (1980).

69 *See* 42 U.S.C. § 1973(b).

70 S. Rep. No. 97-417, pt. 1, at 28–30; *see also* Garza v. County of Los Angeles, 918 F.2d 763, 770 (9th Cir. 1990) (describing the background of Voting Rights legislation).

71 *See* Thornburg v. Gingles, 478 U.S. 30, 50–51 (1986).

72 *See* 42 U.S.C. § 1973(b); *see also* Abrams v. Johnson, 521 U.S. 74, 91 (1997) (stating that "[o]nce plaintiffs establish these conditions, the court considers whether, 'on the totality of circumstances,' minorities have been denied an 'equal opportunity' to 'participate in the political process and to elect representatives of their choice.'").

73 Bush v. Vera, 517 U.S. 952, 959–63 (1996); *see also Abrams*, 521 U.S. at 91 (citing *Bush*).

74 *See Bush*, 517 U.S. at 977; Miller v. Johnson, 515 U.S. 900, 920–23 (1995).

75 *See e.g.*, Mallory v. Ohio, 173 F.3d 377 (6th Cir. 1999), discussed *infra*.

76 *Gingles* applied to multi-member districts; the Supreme Court has since extended the framework to single-member districts. Growe v. Emison, 507 U.S. 25, 40–41 (1993).

may also understand that heteroscedasticity,[77] spatial autocorrelation,[78] or aggregation bias[79] may affect the results and choice of model. But, the judge may not comprehend how the expert's assumptions affect the model, how adding or omitting variables changes the results of the analysis, or even why one model should be used rather than another model. Indeed, if another expert were to testify, reaching different substantive results with different statistical methods, the judge may be unable to meaningfully evaluate the two models. Under these circumstances, the judge ultimately makes a determination on the admissibility of the testimony without having to provide a scientific justification. In addition, *Joiner* ensures that the judge's decision will be well-shielded. Ironically, decisions regarding the admissibility of scientific evidence may be protected as much by institutional design as by *Joiner's* deferential abuse of discretion standard.[80]

### D. EXISTING MEASURES TO ASSIST THE JUDGE

A trial judge, may, if she so chooses, seek assistance from outside her chambers when evaluating scientific evidence. First, pursuant to Rule 706 of the Federal Rules of Evidence, the judge can select a court-appointed expert, i.e. appoint her own scientific expert, with or without the consent of the parties or the knowledge of the jury.[81] The costs are borne by the parties, "in such proportion and as such time as the court directs."[82] Second, the judge can appoint a special master, i.e. empower a third person to conduct proceedings and to make a written report to the court, and to rule on the admissibility of evidence.[83] As with a court-appointed expert, the costs for a special master are borne by the parties in a way determined by the judge. Third, a judge may seek a technical advisor who performs the same role as a court-appointed expert or special master, but purely an advisor, not subject to being deposed or required to testify at trial.[84] The use of experts, special masters, or

---

77 Disturbances or error terms in a regression equation are heteroscedastic when they have different variances. One assumption of the classical normal linear regression model is that the variance of the disturbance is constant for all observations.

78 Spatial autocorrelation occurs when there is spatial dependence arising from geographical location.

79 This condition occurs when the parameters and the regressors in the regression model are correlated.

80 Federal circuit judges are drawn from the same pool as district judges—albeit often the higher end of the distribution—and for this reason are generally no more qualified to review the admissibility decisions of the district judge than the judge is to make the initial determination. The individual and institutional constraints facing a district judge do not differ significantly from circuit judges in matters of statistical understanding.

81 FED R. EVID. 706(a), (b).

82 FED. R. EVID. 706(b).

83 FED. R. CIV. P. 53.

84 FED. R. EVID. 706.

advisors enables judges to seek the assistance that they need but do not possess, either personally or within the chambers, and do not receive from the litigants themselves (which is often the case when one of the litigants lacks the financial resources to hire his own expert to challenge the findings of the other expert).

These approaches, however, have limited usefulness. Both Rule 706 and Rule 53 impose costs that one or both parties may be unable to pay since experts of suitable caliber may be prohibitively expensive. Moreover, use of Rule 53 is favored only as "exception and not the rule."[85] For non-jury cases, the Supreme Court has held that the complexity of issues does not justify use of a special master.[86] Ultimately, the most obvious difficulty is the bias of the expert. Whether intentionally or unwittingly, an expert may fail to provide the judge with a complete picture of the pros and cons of particular scientific evidence, so the judge may simply be reinforcing the views of one of the litigant's experts. Even when experts for both sides testify, the purportedly neutral advising expert who is philosophically opposed to either the scientific approach or the opposing testifying expert may accentuate the limitations of an approach without rightfully acknowledging its probative, and therefore, admissible value.

## III. THE METHODOLOGICAL ISSUES

### A. THE ECOLOGICAL INFERENCE PROBLEM

The conceptual framework for cases involving expert witnesses clearly seems to be problematic on several fronts. To determine precisely how troubling this framework is, however, requires a deeper look into case law to observe how these problems are dealt with in practice. As we previously mentioned, Voting Rights cases almost inevitably involve making cross-level inferences. That is, we have data at some aggregate-level (often at the precinct-level and sometimes at a higher level of aggregation such as a city or a county), but we are interested in the behavior of minorities or some group other than the level of aggregation available to us. Hence, we must infer characteristics of one group based on our observations of a cross-level of data. This is the ecological inference problem.[87] This problem arises in the specific context of Voting Rights cases because of the specific tests outlined in *Thornburg v. Gingles*.[88]

---

85 FED. R. CIV. P. 53(b).

86 La Buy v. Howes Leather Co., 352 U.S. 249, 259 (1957).

87 *See generally* CHRISTOPHER H. ACHEN & W. PHILLIPS SHIVELY, CROSS-LEVEL INFERENCE (1995).

88 478 U.S. 30 (1986).

In mathematical terms, the ecological inference problem is exactly an ill-posed inverse problem. The definition of an ill-posed inverse problem posits that no unique inverse or solution exists. In our context, given the aggregate data, there are many permutations of individual-level preferences which will yield the observed aggregate preferences. Thus, in order to present a single solution, one must impose preferences by introducing a selection criterion. In doing so, one must choose from among a set of desirable attributes for the solution. The nonuniqueness of solutions forces an arbitrary choice if one insists on representing the answer to the problem by a single model. Hence, the insistence on a "solution" may often be unreasonable and unwarranted by the data. Unfortunately, the insistence on one answer frequently remains unavoidable since judges, lawyers, and social scientists alike are often more interested in a "solution" than in the intricate details of the model formulation or the fine art of devising solutions to an ill-posed inverse problem.

*Garza v. County of Los Angeles Board of Supervisors*[89] illustrates these claims. The data used in this case were tract-level data from Los Angeles County. In addition to the number of votes certain candidates received, other tract-level information such as the citizen voting age population, the number of high school graduates, the total registration, and home ownership figures, among others, were available. In general, there are a number of variables which are potentially available for any such analysis. For instance, information that is collected from the census can be merged to their tract-level counterparts.

The additional information that one brings to the model can be pivotal. Though not a very spectacular claim, it is one that is often lost upon those working in the area of Voting Rights. For instance, one source of error occurs at the data collection stage. The *Garza* data include the number of registered Hispanics in each precinct. The count was surmised by matching the registration data to a Hispanic surname dictionary. There are many issues involved in parsing and coding the data correctly. Mistakes inevitably occur.

Another source of bias stems from issues of model specification.[90] For instance, there is considerable uncertainty as to which variables should enter an analysis of aggregate data, and the decision to include some variables and exclude others can affect the results greatly. Ultimately, there is no way to make a "correct" ruling in these situations, whether it is by a judge or a statistician. The problem is that information is lost, and though this information may be recovered by properly specifying a model, one is never certain that any particular specification is the

---

89 918 F.2d 763 (9th Cir. 1990).

90 For an excellent discussion of these issues, *see* CHRISTOPHER H. ACHEN & W. PHILLIPS SHIVELY, CROSS-LEVEL INFERENCE (1995).

correct one. A number of assumptions must be made in order to estimate the parameters. Often, the choice of one assumption over another is somewhat arbitrary in that one cannot claim that a certain choice is more virtuous. In the end, many of these decisions are simply choices with questionable justification. Another specification issue involves the number of groups that one should include in the analysis. Generally, if the voting rights of Hispanics are at issue, the court has claimed that a bivariate regression, where percent Hispanic is the only independent variable, is adequate. However, there is no justification for this specification if, say, the percentage of any other minority group is an issue in the outcome of the election. From what we know substantively, this is almost always the case.

It is simple to illustrate that the issues with specification are critical. We demonstrate this point by re-analyzing the original data from *Garza*. Here, we compare the results from several different statistical methods: homogeneous precinct analysis, the correlation coefficient, the neighborhood model, ordinary least squares ["OLS"],[91] and EI (a random coefficients model proposed by Gary King).[92] In more detail, we will examine various specifications for OLS and EI. These two methods warrant a closer look since both of these methods have been used in Voting Rights cases. OLS was used in two landmark Voting Rights cases, *Garza* and *Gingles*, while EI was the choice method in a more recent case, *Mallory v. Ohio*.[93] Although the judge in *Mallory* was effusive about the advances in the EI model, we will see that OLS and EI are *extremely* similar. Rarely do they offer results that differ significantly. King's statement "[u]nfortunately, even the best available current methods of ecological inference [i.e., OLS] are often wildly inaccurate,"[94] is ironic in this respect and is essentially a statement against his own proposed

---

[91] OLS is a statistical technique for estimating a linear relationship between a set of explanatory variables and a response variable. In the context of Voting Rights cases, the response variable is usually the proportion of the vote obtained by a certain candidate. The set of explanatory variables often includes only a measure of racial proportion. Some refer to OLS as "Goodman's regression." Apparently, this is a reference to Goodman's 1953 article in which he expounded on the pitfalls of using OLS in aggregate data analysis. However, it is extremely clear from both that article and Goodman's 1953 article on the same topic that he never advocated widespread use of OLS with aggregate data. In fact, he cautioned against it, stating that it was appropriate only "under very special circumstances." Hence, it is unfortunate that anyone would attribute the name "Goodman's regression" to the use of OLS on aggregate data. Goodman certainly would not have used OLS in the slipshod manner that has become commonplace today. *See* Leo A. Goodman, *Ecological Regressions and Behavior of Individuals*, 18 AM. SOC. REV. 663–64 (1953).

[92] *See generally* GARY KING, A SOLUTION TO THE ECOLOGICAL INFERENCE PROBLEM: RECONSTRUCTING INDIVIDUAL BEHAVIOR FROM AGGREGATE DATA (1997). As discussed *infra*, Dr. King is considered a prominent statistician in academic and legal communities.

[93] 38 F. Supp.2d 525, 538 (S.D. Ohio 1997), *aff'd*, 173 F.3d 377 (6th Cir. 1999).

[94] KING, *supra* note 92, at 15.

method. At best, EI offers a modest improvement over OLS and is, in fact, inferior in some respects.

### B. EVIDENCE FROM VARIOUS TECHNIQUES

In *Garza*, Hispanic residents of Los Angeles County sought a redrawing of the district boundaries for the county's Board of Supervisors, alleging that the existing boundaries were drawn so as to dilute Hispanic voting strength.[95] Specifically, under Section 2 of the Voting Rights Act, they asked that the district lines be redrawn to create a Hispanic-majority district.[96] The district court found for the plaintiffs, holding that the County deliberately diluted the Hispanic Vote, thereby violating the Voting Rights Act. In making its determination, the court relied heavily on expert witnesses.[97]

Homogeneous precinct analysis is a technique that is advocated by Bernard Grofman and was used in both the *Garza* and *Gingles* cases.[98] The basic idea behind the technique is that we have perfect information about how a group voted if it comprises an entire precinct.[99] While this is rarely the case, other precincts may be nearly homogeneous in this way. By Grofman's standards, a precinct is homogeneous if over 90% of its registrants are from one group.[100] Near homogeneity allows one to be extremely confident when making claims about group behavior. In the *Garza* data, there are thirteen precincts in which registered Hispanics account for over 90% of the registered voters. Of these thirteen precincts, the average Hispanic support for the Hispanic candidate is 38.1% with a standard deviation of 2.6. The minimum support is 33.1%, while the maximum support is 41.6. Since there is not a great deal of variation, these numbers imply that Hispanic support for this candidate is relatively low and somewhat apathetic. This conclusion is further bolstered by the substantive belief that Hispanics in more heterogeneous districts are likely to support Hispanic candidates at lower rates than those who reside in homogeneous precincts. Generally, we expect homogeneous precincts to display the highest levels of support.

Examining the Hispanic support rate is informative. However, to determine whether there is racial polarization, that is, whether the groups vote in opposing manners, we must compute non-Hispanic support rates as well. There are a large number of non-Hispanic precincts in Los Angeles, 938 in total. Conducting this same analysis on these nearly homo-

---

95 918 F.2d at 766.

96 *Id.* at 765 (describing the plaintiff's argument).

97 *See id.* at 767 (outlining the statistical evidence upon which it relied).

98 *See* BERNARD GROFMAN ET AL., MINORITY REPRESENTATION AND THE QUEST FOR VOTING EQUALITY 88–90 (1992).

99 *See id.* at 84–85.

100 *Id.* at 85.

geneous precincts for non-Hispanics yields a support rate of 18.4%. This non-Hispanic support rate is significantly lower than the Hispanic support rate. However, it still would be difficult to argue that there was racially polarized voting since neither group seemed to support the candidate.

Another method of assessing whether the vote for a Hispanic candidate rises with increased Hispanic registration is to look at the correlation coefficient. For these data, the correlation coefficient is 0.61, implying that as support for the Hispanic candidate increases, the proportion of Hispanics registrants tends to rise as well. However, the relationship can be construed only as a weak general pattern since 0.61 is a fairly moderate level of correlation. Even if the correlation coefficient were higher, it may not be an accurate measure of the underlying individual behavior. This has been well-known since 1950.[101] Lupia and McCue document that the correlation coefficient can be exactly the same for two different electorates who vote in very different manners.[102] Hence, the evidence, if any exists at all, can be construed as weak, at best. Given that, the correlation coefficient has been combined with homogeneous precinct analysis as a source of additional evidence. In this case, one might say that the relationship is generally increasing and peaks at a support rate around 40%. This would provide stronger evidence that the degree of racially polarized voting was not very high.

We cannot make very strong claims based on either homogeneous precinct analysis or the correlation coefficient. Certainly part of not being able to make very strong claims is a result of not imposing strong assumptions. In order to make stronger claims, that is, to make statements about the actual rate at which Hispanics voted for the Hispanic candidate requires the imposition of a set of assumptions, which may or may not be true. While some assumptions may seem more credible than other assumptions, one cannot be certain which assumptions hold. The problem is not with formulating a model where the parameters can be estimated. The problem is that we have no idea whether any of these formulations incorporates appropriate assumptions or produces accurate estimates of the underlying individual-level behavior.

One method that is widely criticized as imposing unbelievable assumptions is the neighborhood model advocated by Freedman in the *Garza* case.[103] In this model, the assumption is that voters tend to vote alike if they are from the same neighborhood, and there is no systematic

---

101 *See* W. S. Robinson, *Ecological Correlations and the Behavior of Individuals*, 15 AM. SOC. REV. 351 (1950); *see also* Arthur Lupia & Kenneth McCue, *Why the 1980s Measures of Racially Polarized Voting Are Inadequate for the 1990s*, 12 L. & POL'Y 353, 355–64 (1990).

102 *See* Lupia & McCue, *supra* note 101, at 355–64.

103 *See* KING, *supra* note 92, at 194.

difference in voting tendencies between groups. Instead, people who live near each other will have similar statuses as well as political opinions.[104] The variation that exists is manifested from precinct to precinct. Though this assumption is widely regarded as untenable, there is substantive and empirical support for neighborhood effects in the political science literature,[105] and it has been validated on some data sets where the answer is known.[106] The downfall is that this assumption is generally stronger than the social context effects discussed in the literature. In other words, there are neighborhood effects, but the neighborhood effects are not so pronounced that groups would display the exact same behavior. Race often remains a source of additional heterogeneity. If the neighborhood effect is not the predominant effect, the neighborhood model will not likely produce estimates that are close to the truth.

For our data, the neighborhood model's estimate of Hispanic support is 26.4% while the non-Hispanic support rate is 23.9%. Again, both of these support rates are relatively low and relatively close to one another. These two situations together again imply that no significant polarization exists. So, the neighborhood model's estimates could be considered additional evidence for the lack of polarization since they are consistent with the estimates from homogeneous precinct analysis and the correlation coefficient. In general, the neighborhood model is predisposed to produce results that imply no polarization. The underlying assumption, after all, is that voters who live in the same neighborhood vote alike regardless of race. Hence, evidence of polarization from the neighborhood model should be given the emphasis it deserves and lack of evidence should be cautiously viewed since that is the predisposition.

The model of choice in Voting Rights cases, or the choice of judges at any rate, has so far been OLS or ecological regression. The main point of contention with the OLS model is that the model incorporates the very strong "constancy assumption," that is, outside of random variation, a group tends to vote for a certain candidate in the same proportions, regardless of precinct of residence. Unlike the neighborhood model that assumes that neighborhood effects are the paramount and, indeed, only factor in determining vote choice, this assumption posits that race is essentially the *only* factor in voting decisions. Certainly, both of these assumptions are very strong and neither is likely to be de facto true. More likely, there will be at least some neighborhood effect *and* some race effect. Blacks who live in affluent, predominantly white neighborhoods

---

104 *See* David Freedman et al., *Ecological Regression and Voting Rights*, 15 EVALUATION REV. 673, 682–87 (1991).

105 *See, e.g.*, BERNARD R. BERELSON ET AL., VOTING (1954); HEINZ EULAU, POLITICS, SELF, AND SOCIETY (1986); Michael MacKuen & Courtney Brown, *Political Context and Attitude Change*, 81 AM. POL. SCI. REV. 471, 478–85 (1987).

106 *See* Freedman, *supra* note 104, at 687–94.

have been documented to have different preferences and voting behavior than blacks who live in the inner city.[107] Likewise, race has been shown to be politically divisive.

The OLS estimate of Hispanic support is 47.6% while the estimate of non-Hispanic support is 17.2%.[108] These support rates differ significantly from those that were estimated from the alternative models. Of particular interest is the disparity between support rates; the other models predicted the support to be similar. This model implies that there is racial polarization. Certainly, this inconsistency raises some concerns. Moreover, a method that is generally regarded as better for these data, weighted least squares ["WLS"], yields similar results, 48.5% and 17.2% respectively.[109]

One might alternatively consider utilizing a random coefficient model. The EI model is one particular specification of the broad class of random coefficient models. We examine this particular specification because it has received widespread publicity, specifically, articles in the *New York Times* and *Boston Globe*, and has been used in Voting Rights cases.[110] The assumptions of EI, contrary to much of the publicity espousing its virtues, are at least as strong as those of OLS. In fact, EI imposes all the assumptions of OLS and more.[111] EI does not impose the constancy assumption but does impose an analogous "similarity assumption." In particular, EI imposes the assumption that while a group does not vote for a candidate in the *same* proportions, it does vote for a candidate in "roughly the same" proportions. In fact, the similarity assumption is so close to the constancy assumption that EI and OLS produce indistinguishable estimates in the vast majority of cases.

Figure 1 displays a Monte Carlo experiment with 1000 data sets (2000 parameters estimates include 1000 estimates for each parameter). Data were generated, aggregated, then estimated with both EI and OLS. As we can see, the EI and OLS estimates are very similar. This is obvious by noting that most of the points lie along the 45 degree line. In fact,

---

107 *See* JEFF MANZA & CLEM BROOKS, SOCIAL CLEAVAGES AND POLITICAL CHANGE: VOTING ALIGNMENTS AND U.S. PARTY COALITIONS, 160–61 (1999).

108 Note that our estimates are not a perfect replication of the results reported in the *Garza* case. The estimates are close however. The discrepancy arises from a slight difference in the data. The actual data used in the *Garza* case were precinct-level data while our data are tract-level data. Hence, the unit of aggregation differs. The differences are slight and generally within one percentage point.

109 In general, WLS addresses issues of efficiency rather than bias, so the similarity here is not shocking.

110 *See* Karen Freeman, *Statistician Builds What May Be a Better Data Mousetrap*, N.Y. TIMES, July 8, 1997, at C8; John Yemma, *Connecting the Dots of Data*, BOSTON GLOBE, July 27, 1997, at 5.

111 *See* Douglas Rivers, *Review of A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, 92 AM. POL. SCI. REV. 442 (1998) (book review) (citing both the attributes and limitations of King's contribution).

the correlation between the parameter estimates from the two models is 0.98—high by *any* standards. There are a few estimates which seem to be outliers. Some of these outliers are simply the result of poor and buggy code in the EI computer program. Independent sources have documented EI's frequent erratic computations.[112] The other differences occur when the estimates lie near the [0,1] bounds.

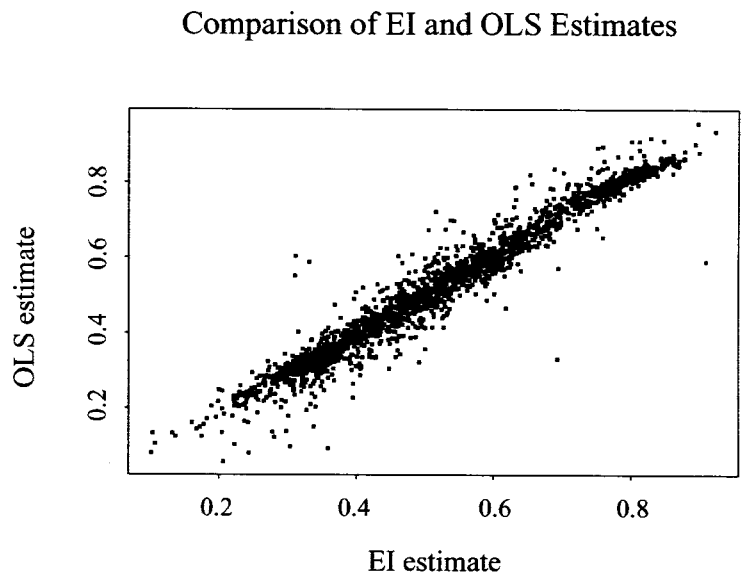



FIGURE 1: The correlation coefficient resulting from over 2000 different EI and OLS estimates is 0.98. The vast majority of these points lie exactly on the 45 degree line. Few points deviate from this clear pattern.

For the *Garza* data, as expected, the EI estimates are close to the OLS estimates. The EI estimate of Hispanic support is 45.9%, and the estimate of non-Hispanic support is 17.6%. Given this evidence, the novelty of the EI estimator is overstated given its striking similarity to OLS. Moreover, the added complexity and small gain in efficiency are hard to justify given that they are coupled with problematic software. There are occasions when EI will give estimates inside the [0, 1] bounds

112 *See, e.g.*, David Freedman et al., *Review of A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, 93 J. AM. STAT. ASS'N. 1518, 1518–20 (1998) (book review); Wendy K. Tam Cho & Brian J. Gaines, *Reassessing the Study of Split-Ticket Voting*, AM. POL. SCI. REV. (forthcoming, 2001). Freedman independently coded the procedure described in King's book and arrived at different estimates. Cho and Gaines document problems with EI's sampling procedures. In particular, they show that multiple runs of the EI program on the same data set often yield inconsistent answers. The estimates from each run can differ significantly. Indeed, King himself notes many bugs in the documentation of his software updates. Insinuations that bugs do not exist are thus disingenuous and simply attempts to hide problems with the software.

when OLS produces out-of-bounds estimates. These seem like clear instances when EI would be a better estimator than OLS. However, while EI gives estimates within the bounds, it does not necessarily give correct estimates. We are not interested in *possible* estimates but in *correct* estimates. EI does *not* bridge this chasm. Indeed, if the specification is incorrect, the estimates are biased and inconsistent. With out-of-bounds OLS estimates, one is given the luxury of knowing that the model is misspecified. EI masks this misspecification by giving an estimate within the bounds. Hence, this "feature" is not a feature at all if the estimates are not close to the truth.

At issue then is the ability to find the correct specification for a given data set. Substantively, we have documented evidence that group voting rates are not constant when other variables such as partisanship, level of income, or education change. Few would claim that wealthy, Republican Hispanics living in affluent areas display the same voting tendencies as poor, Democratic Hispanics living in the inner city (the OLS constancy assumption). Moreover, few would assume that their voting preferences are similar (the EI similarity assumption). In both EI and OLS, we are able to take other variables into consideration. The problem is not whether we can consider other variables but whether these other variables really affect the vote or are merely erroneously thought to influence the vote.

The issue with choosing covariates is choosing proper covariates. Since the information lost in the aggregation is irretrievable, choosing covariates based on substantive knowledge is not a good method and obviously can result in estimates that are very far from the truth. The choice of covariates is at least as arbitrary as choosing the type of model (i.e. neighborhood model, OLS, or EI). There are some methods designed for choosing proper covariates in aggregate data.[113] Certainly, it is better to have systematic techniques than to choose covariates without a method. Freedman et al. discuss the use of multiple regression as a viable alternative if one were able to find a proper specification. The problems they state are "Which variables should go in? How should they be measured? Should the regression be linear, logistic, or something else?"[114] At least some of these issues are the covariate selection issue. Some progress was made in this area since their article.[115]

---

113 *See* Wendy K. Tam Cho, *Latent Groups and Cross-Level Inferences*, 20 ELEC. STUD. 243–63 (2001); Wendy K. Tam, *Structural Shifts and Deterministic Regime Switching in Aggregate Data Analysis* (1997) (Master's Essay, Dep't of Statistics, U.C. Berkeley).

114 Freedman, *supra* note 104, at 698.

115 *See* Cho, *supra* note, 113.

## IV. DISCUSSION, POLICY PROPOSALS, AND RECOMMENDATIONS

In Voting Rights cases, the choice of statistical methodology is the critical decision. The decision of which methodology to accept, determined solely by the judge, often determines whether the plaintiff or defendant wins. If the judge is unable to make an independent assessment of a model, he may have to rely on the judgment of others in the academic community. This decision by the judge, because of *Joiner* and the limited technical knowledge of appellate judges, will be well-shielded.

In *Mallory v. Ohio*, the Sixth Circuit affirmed the district court decision that voting rights were not violated. The trial judge could have brought in an additional expert to evaluate Dr. King's testimony on behalf of the defendants in the judicial elections of eight of Ohio's largest districts, but elected against it, relying on Dr. King's credentials and the supporting expert testimony of two political scientists lauding Dr. King's work.[116] Upon review, the Sixth Circuit merely echoed the findings of the trial judge, stating that Dr. King is "the world's foremost authority on the statistical analysis of racial bloc voting . . . [whose] techniques have been widely recognized in the academic community and utilized in other voting rights cases."[117] Like the trial court, the Sixth Circuit noted, but did not find troubling, that the plaintiffs "presented no statistical evidence to support the[ir] claims."[118] A balance of expert testimony provides both the district and appellate court greater understanding of the strengths and limitations of the testimony.[119]

In Voting Rights cases as well as in other cases that involve statistical models, the bottom line lies with the assumptions made in the model. Although we need to make assumptions, we have no idea whether the assumptions we make are true. This uncertainty must be acknowledged. Every time we change the assumptions, we change the estimates, and often the outcome. If no consensus exists in the estimates from different models, then the Court should be especially skeptical.

The court should not use this lack of consensus—assuming that the judges are aware one exists—as an excuse to choose the methodology that best suits a judge's desired outcome. Clearly, by those well-versed

---

116 *See* 38 F. Supp.2d 525, 540. While political scientists certainly engage in the analysis of data, it seems odd that one would consult social scientists on this matter rather than tapping the advice of statisticians. Perhaps this decision was intentional on the part of the defendants, given that Dr. King's method is not established and accepted by the statistical community. *See* Freedman *supra* note 104.

117 173 F.3d at 383.

118 *Id.*

119 *See, e.g.*, Cousin v. Sundquist, 145 F.3d 818, 826 (6th Cir. 1998), a Section 2 Voting Rights Case in which the Sixth Circuit reversed the district court after reviewing the experts' testimony for both the plaintiff and defendant.

in the art, the outcome of a case is highly manipulable. Judges often write as if they are aware of the complexities, stating with conviction that one method is the proper method. However, they may be unaware.[120] The threat to voting rights litigants—plaintiffs and defendants alike—is that court decisions will find a voting rights violation when none occurred or vice versa. If decisions are made on the basis of subjective ideological biases, they should not be masked under the guise of proper science.

There are two main proposals for dealing with the scientific evidence challenge facing courts. The first is the creation of a panel of experts for each particular field, drawn from the academic or private sector community.[121] One or more members of the panel would provide assistance to federal judges in need, and would aver that they have no personal interest in the outcome of the case. Conceptually, this proposal is persuasive because it reduces the bias of the expert and makes it easier for judges to seek assistance. As a practical matter, however, it appears unlikely that experts would collectively commit to such a body. Many of the top experts, the precise group from which judges would like to solicit advice, would rather spend their time elsewhere, either by providing their expert consulting fee[122] or simply to work on their own research. The issue of expert fees concerns courts,[123] and some courts have sought to reduce fees of testifying experts. Experts, of course, are free not to lend their services for a lower fee.

The second approach is to create a specialty court to deal specifically with scientific issues. This court would be analogous to other existing Article III specialized federal courts: bankruptcy, tax, and patent law. Cases which require an understanding of complex scientific issues could be referred to these courts. The problem is that, at least in the area of voting rights, specialty courts seem to be a nonviable solution. There

---

120 *See, e.g.*, the district court's statement, "Dr. King's methodology constitutes an improvement upon the 'Goodman's regression' method of analysis that was used by the experts in *Gingles* and ultimately relied upon by the United States Supreme Court." *Mallory v. Ohio*, 38 F. Supp.2d at 538. The judge in this case clearly does not understand the complexities that would be involved in a proof of the superiority of King's ecological inference estimator. In fact, the limitations of King's estimator are many, and it does not clearly constitute an improvement over the classical regression approach. *See* Wendy K. Tam Cho, *Iff the Assumption Fits. . . : A Comment on the King Ecological Inference Solution*, 7 POL. ANALYSIS 143–63 (1998).

121 *See* Ellen E. Deason, *Court-Appointed Expert Witnesses: Scientific Positivism Meets Bias and Deference*, 77 OR. L. REV. 59, 74 (1998).

122 *See* James A. Mellowitz, *Whatever the Market Will Bear: Fighting Exorbitant Expert Fees with Rule 26(B)(4)(C)(I)*, 38 RES GESTAE 15 (1995) (giving examples of exorbitant fees).

123 *See, e.g.*, Jochims v. Isuzu Motors, Ltd., 141 F.R.D. 493, 497 (S.D. Iowa 1992) ("Continuing escalation of expert witness fees . . . is of great concern. The escalating cost of civil litigation runs the grave risk of placing redress in the federal courts beyond the reach of all but the most affluent.").

are not enough voting rights cases across the country annually to justify creating a specialty court.

Since neither proposal appears likely to gain acceptance in the near future, it appears as though judges will continue to evaluate scientific evidence as he or she sees fit, subject to an abuse of discretion standard. The current system gives trial judges considerable discretion to be ambitious or lazy regarding scientific evidence since the judges are never required to justify, at least on rigorous scientific grounds, why they chose to exclude or admit evidence. Under an abuse of discretion standard, the reviewing courts are also not required to research deeply why the evidence was excluded.

The remaining option, therefore, is for judges to use the existing structure more effectively. For example, when a plaintiff seeks to have an expert testify that a voting rights violation occurred based on statistical inference, the judge can, and should, require that another expert testify to challenge (or at least comment on any limitations of) the testimony of the first expert.[124] This second expert ideally will be provided by the other defendant, or if need be, the judge can appoint his own expert, pursuant to Rule 706, and impose some or all of the costs to the party who initially failed to provide its own expert. If cost remains a central issue, the judge can require that both parties bear some of the cost. The plaintiff may dislike having to subsidize the cost for a defense expert, and must therefore engage in a benefit-cost analysis to decide whether the testimony of its own expert is worth the additional cost. Collectively, if judges establish that they will require, as a matter of form, a second competing (or qualifying) expert in matters of scientific evidence, this should provide the added benefit of motivating plaintiffs and defendants alike to present testimony that is within the bounds of acceptability, even if not within the bounds of "general acceptability." The Supreme Court could impose this procedural requirement on trial courts simply by stating that whenever expert testimony is used, it must come from both the plaintiff and defendant. Or, alternatively, appellate courts could invoke their supervisory powers to also require this of trial courts.[125]

If the judge is still unable or feels inadequate to adjudicate between dueling consultants/experts but must make the technical call, the judge may employ a model averaging approach. That is, the judge could re-

---

124 Of course, the same principles apply when the defendant rather than plaintiff offers an expert to testify that a voting rights violation has not occurred.

125 *See, e.g.*, United States v. McDowell, 814 F.2d 245, 249–50 (6th Cir. 1987); United States v. Bailey, 675 F.2d 1292, 1297 (D.C. Cir. 1982), *cert. denied*, 459 U.S. 853, 103 S.Ct. 119 (1982) (both courts invoking their supervisory powers to identify the nature of the inquiry to be made and the procedure to be followed henceforth in situations where an accused seeks to waive representation by counsel and proceed *pro se*).

quire both experts to provide the results from estimating a set of models that they deem adequate or acceptable. At this point, if the results from various models are relatively stable, the judge might then feel confident in making a ruling based on these consistent estimates from differing statistical analysis. If the results do differ, the judge may consider that there is uncertainty in the different specifications. To account for or mitigate the uncertainty, the judge might average the results from the different model specifications. This approach leaves far less to the discretion of ideologically-inclined and statistically-untrained judges. Bayesian statisticians have proposed coherent procedures for model averaging.[126] Hence, instead of treating the results from a particular model as an answer with any certainty, the judge can incorporate the uncertainty and disagreement into the final result under a meaningful statistical framework.

In summary, while the present structure for dealing with technical information is not ideal, judges do have some means for overcoming at least some of the information hurdles posed by scientific evidence, and in the process, align the incentives of plaintiffs and litigants to present "acceptable" testimony. Some of these avenues involve changing the structure under which technical cases are decided. Some involve working within the constraints of the structure. All provide a means by which to transform a problematic situation into one that conforms more closely to ideals of the legal system. The unanswered and more significant question is whether the judges have the incentive to take this initiative.

---

126 *See* Larry M. Bartels, *Specification Uncertainty and Model Averaging*, 41 AM. J. POL. SCI. 641–74 (1997); David Draper, *Assessment and Propagation of Model Uncertainty*, 57 J. ROYAL STAT. SOC. SERIES B 45–97 (1995); Adrian E. Raftery et. al, *Bayesian Model Averaging for Linear Regression Models*, 92 J. AM. STAT. ASS'N. 179–91 (1997).