ARTICLES

COOPERATION AND DIVISION: AN EMPIRICAL ANALYSIS OF VOTING SIMILARITIES AND DIFFERENCES DURING THE STABLE REHNQUIST COURT ERA— 1994 TO 2005

*Mark Klock**

The Stable Rehnquist Court Era (SRCE) covers the period from the appointment of Justice Breyer to the passing of Chief Justice Rehnquist. There has been only one longer period of stability in the Court's history, and that was in the early nineteenth century when far fewer cases were decided. Because the composition of the Court held constant for so long, the SRCE presents a unique opportunity to conduct a statistical analysis of the Justices' votes. I present a statistical empirical analysis of voting for this period, both for the potentially interesting results and as an example of how to conduct and present an empirical study which is objective and replicable. Some of the findings include the following: only a few pairs of Justices have statistically significant differences in voting records; the magnitude of the departure from independent voting is enormous in statistical terms; Justice Thomas is the most predictable Justice; and Justice Scalia is the least-changed Justice. Of particular interest is a finding that is contrary to conventional wisdom. Conventional wisdom suggests that the median Justice closest to the center, presumably Justice Kennedy, is the most influential Justice. However, I have developed a measure of influence which employs the statistically significant effects the Justices have on each other, and this suggests that the most influential Justices on the Court during the SRCE were Rehnquist, Souter, and Breyer.

* B.A., The Pennsylvania State University, 1978; Ph.D. in Economics, Boston College, 1983; J.D. (with honors), University of Maryland, 1988; admitted to the Maryland Bar, 1988; admitted to the District of Columbia Bar, 1989; member of the Executive Board, Center for Law, Economics, and Finance, The George Washington University School of Law; Professor of Finance, The George Washington University School of Business, Washington, D.C.

INTRODUCTION

For more than a decade, a sizeable and respectable community of legal scholars have sought to bring a more rigorous and scientific type of approach into legal research.[1] Undoubtedly, this is motivated in part by a feeling that legal scholarship has softer standards than other social science scholarship that requires such scholars to follow a more scientific approach—developing and using mathematical models and statistical methods—in order to achieve promotion and tenure.[2] Most legal scholars do not work with data or statistics and do not test hypotheses; hence, their results are not subject to the same level of peer scrutiny as those of social scientists, since there are no results to scrutinize.[3] Most legal scholarship consists of persuasive arguments using legal authority and sometimes empirical results borrowed from other fields.[4]

This Article adds to the growing volume of empirical legal literature with an analysis of the voting behavior of the nine United States Supreme Court Justices during what I refer to as the Stable Rehnquist Court Era (SRCE), which began with the appointment of Justice Breyer and

---

[1] *See, e.g.*, Gregory Mitchell, *Empirical Legal Scholarship as Scientific Dialogue*, 83 N.C. L. REV. 167, 168 (2004) (citing several prominent legal scholars calling for more empirical and scientific legal research); Thomas S. Ulen, *A Nobel Prize in Legal Science: Theory, Empirical Work, and the Scientific Method in the Study of Law*, 2002 U. ILL. L. REV. 875, 876 (2002) (advancing the argument that legal scholarship has become more scientific).

[2] *See* Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 125–26 (2002) (observing that other disciplines do not count non-peer reviewed articles towards promotion and tenure and therefore do not take legal scholarship seriously).

[3] *See id.* at 9 (suggesting that most law professors do not receive adequate training in statistical inference and hypothesis testing).

[4] *Cf.* David M. Flores et al., *Examining the Effects of the Daubert Trilogy on Expert Evidence Practices in Federal Civil Court: An Empirical Analysis*, 34 S. ILL. U. L.J. 533, 536 (2010) ("While *Daubert* and its progeny inspired an abundance of literature in the legal community, only a small proportion of this work represents systematic research directly examining the effects of the changes in admissibility standards.") (footnote omitted).

concluded with the passing of Justice Rehnquist.[5] In an effort to make the approach more like other social sciences, I collected data, established a methodology for analysis, and I now report the results for others to scrutinize and either replicate or dispute.[6] The analysis contains some complex models and hypothesis tests, but it also contains simple empirical measurements that are informative.

Some of the results are not surprising. For example, four of the Justices' votes have a negative correlation with conservative outcomes, two have a weak positive correlation with conservative outcomes, and three have a strong positive correlation with conservative outcomes.[7] Which Justices fall into which category will not come as a surprise to legal scholars.[8] Nevertheless, there is value in having a meaningful quantitative measure of the magnitude of conservatism which confirms our rough judgments.[9] There is also value in knowing the precise magnitudes of certain variables even if our knowledge of relative rankings is not shaken.[10] For example, it will be no surprise to scholars of the Court that Justice Stevens is the most contrarian Justice measured by frequency of sole dissents.[11] But it might not be widely known that Justice Stevens

---

5 *See generally* Michael J. Gerhardt, *The New Religion*, 40 CREIGHTON L. REV. 399, 399 (2007) (stating that when Justice O'Connor retired in 2006, Justice Breyer was just "a couple months shy of the record for the longest serving junior justice in American history.").

6 The data is contained in an Excel spreadsheet which I will make available upon request following publication. The statistical analysis was conducted using an econometric statistical package called Shazam, version 7.0. *See generally* KENNETH J. WHITE, SHAZAM: THE ECONOMETRICS COMPUTER PROGRAM VERSION 7.0 USER'S REFERENCE MANUAL 1–71, 255–60, 317–22 (1993) (covering features of the computer program used in generating this study).

7 *See infra* Table 2.

8 *See, e.g.*, Theodore W. Ruger et al., *The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking*, 104 COLUM. L. REV. 1150, 1155, 1174–75 (2004) (classifying Justices Stevens, Souter, Ginsburg, and Breyer as more liberal; Justice Kennedy and O'Connor as moderate; and Rehnquist, Scalia, and Thomas as conservative).

9 *See* Lee Epstein & Jeffrey A. Segal, *Trumping the First Amendment?*, 21 WASH. U. J.L. & POL'Y 81, 97 (2006) (discussing the need for "the most precise measure possible" to determine how ideology drives the votes of the Justices).

10 *See generally* Mark Klock, *Finding Random Coincidences While Searching for the Holy Writ of Truth: Specification Searches in Law and Public Policy or* Cum Hoc Ergo Propter Hoc*?*, 2001 WIS. L. REV. 1007, 1015–22 (2001) (explaining that a descriptive statistic is really an estimate of a population parameter which is only useful if we have information about the magnitude of the sampling error) [hereinafter Klock, *Finding Random Coincidences*].

11 *See* Jeff Bleich et al., *Justice John Paul Stevens: A Maverick, Liberal, Libertarian, Conservative Statesman on the Court*, OREGON STATE BAR BULLETIN (Oct. 2007), http://www.osbar.org/publications/bulletin/07oct/stevens.html ("Justice Stevens frequently demonstrates a penchant for sole concurrences and dissents."); *John Paul Stevens*, OYEZ PROJECT AT CHICAGO-KENT COLLEGE OF LAW, http://www.oyez.org/justices/john_paul_stevens ("Stevens' individualistic personality keeps him permanently outside the mainstream of the Court and . . . he lacks the characteristics of a coalition-builder.") (last visited Oct. 19, 2012).

had more sole dissents during this period than the other eight Justices combined, or that he had more than five times as many sole dissents as the second most contrarian Justice—Justice Thomas.[12] By comparison, Justice O'Connor was the sole dissenter in just a single decision during the SRCE.[13] Another finding that is not surprising is that the two Justices most frequently known as the swing votes in close decisions[14] have the highest batting averages, defined as voting with the majority.[15] Having a precise quantitative measure, however, enables us to have a better understanding as to how important the votes of Justices O'Connor and Kennedy were relative to the other members of the Court during the SRCE.[16] Indeed, some of the empirical models suggest that when we examine the influence the Justices have on one another, the two swing vote Justices do not have the most influence on the Court, contradicting some of the conventional wisdom.[17]

Part I of this Article discusses the power and limits of empirical analysis. Part II describes three alternative models of voting: independent, cooperative, and vindictive. Part III describes the data collection process and provides some background into statistical methodology. Finally, Part IV presents the empirical results.

I. POWER AND LIMITS OF EMPIRICAL ANALYSIS

A. *The Power of Empirics and Our Thirst for Facts*

Empirical facts are difficult to dispute.[18] For example, the fact that Ted Williams was the last player to bat .400 during a major league sea-

[12] *See infra* Table 2 and accompanying text (1.25% of 80 sole dissents is one).

[13] *Id.*

[14] *See* Theodore W. Ruger et al., *supra* note 8, at 1184 (indicating that as moderates, O'Connor and Kennedy are often the swing votes); *see also* C. Lincoln Combs, Note, *A Curious Choice:* Hibbs v. Winn *as a Case Study of Justice Sandra Day O'Connor's Balancing Jurisprudence*, 37 ARIZ. ST. L.J. 183, 197 (2005) (referencing Justice O'Connor's "status as the swing vote between the conservative and liberal voting blocs on the Court . . . "); Jennifer S. Hendricks, *Converging Trajectories: Interest Convergence, Justice Kennedy, and Jeannie Suk's "The Trajectory of Trauma,"* 110 COLUM. L. REV. SIDEBAR 63, 65 (2010) (observing "Justice Kennedy's status as the swing vote.").

[15] *See infra* Table 2.

[16] *See infra* Table 4 (showing that Justice Kennedy's voting percentage is not significantly different from Justice Rehnquist's or Justice Souter's statistically, and that Justice O'Connor's voting percentage is not statistically significantly different from the aforementioned Justices and Justice Breyer).

[17] *See infra* Table 7 and the accompanying discussion (concluding that the Justices with the most positive net influence on the Court were Justices Rehnquist, Souter, and Breyer). The conventional wisdom is that swing voters are the most powerful. *See* Lee Epstein & Tonja Jacobi, *Super Medians*, 61 STAN. L. REV. 37, 40 (2008) ("[I]n theory the median Justice should be quite powerful . . . .").

[18] By definition, a fact is something known to be true, "something that has actual existence." MERRIAM-WEBSTER, *available at* http://www.merriam-webster.com/dictionary/fact (last visited Oct. 24, 2011).

son cannot be disputed.[19] Legal decisions, however, are disputable.[20] This explains not only the existence of appeals, but also the need to have a terminal court so that all disputes eventually come to an end.[21] The power of empirical measures lies in their perceived ability to resolve disputes.[22] As constraints on financial and political resources have become more intense, the stakes in policy debates have become greater.[23] The forces of supply (inexpensive computational power and data storage)[24] and demand (the strong desire to have the most persuasive facts available to win a high stakes debate) have worked to bring the power of empirics into battle.[25]

The growth in empirical legal studies has been explosive.[26] In 2002, Professors Epstein and King claimed that empirical research had become commonplace among legal scholars during the previous two decades.[27] To support their empirical observation they reported that 231 papers with the word "empirical" in the title were published in American

---

[19] Fred Bowen, *.400: A Disappearing Magic Number*, WASH. POST, June 27, 2008, at C12.

[20] *See, e.g.*, ALAN M. DERSHOWITZ, SUPREME INJUSTICE: HOW THE HIGH COURT HIJACKED ELECTION 2000 at 81 (2001) ("The majority per curiam opinion [in *Bush v. Gore*] is likely to become one of the most analyzed, criticized, and defended opinions in the history of the Supreme Court.").

[21] *Cf.* Mark Klock, *Is it the "Will of the People" or a Broken Arrow? Collective Preferences, Out-of-the-Money Options,* Bush v. Gore, *and Arguments for Quashing Post-Balloting Litigation Absent Specific Allegations of Fraud*, 57 U. MIAMI L. REV. 1, 17 (2002) (observing that there is value in decisively and conclusively terminating endless disputes) [hereinafter Klock, *Bush v. Gore*].

[22] *Cf.* MICHAEL O. FINKELSTEIN & BRUCE LEVIN, STATISTICS FOR LAWYERS x (2d ed. 2001) ("[I]t seems inevitable that studies based on data will continue to be pursued by the scholarly community and presented with increasing frequency in litigated matters involving public issues.").

[23] *See, e.g.*, Laura K. Abel, *Evidence-Based Access to Justice*, 13 U. PA. J.L. & SOC. CHANGE 295, 310 (2009) ("It is critically important that resource allocation decisions . . . are made accurately because the stakes are high.").

[24] *See* FINKELSTEIN & LEVIN, *supra* note 22, at x ("What is true of the general world has filtered into the courtroom. . . . [T]he ubiquity of data and computers, and the current fashion, have encouraged the creation of elaborate econometric models that are sufficiently plausible to be accepted in learned journals.").

[25] *See* ROBERT COOTER & THOMAS ULEN, LAW & ECONOMICS 9 (6th ed. 2012) ("From economists, lawyers can learn quantitative reasoning for making theories and doing empirical research.").

[26] *See* Corey Rayburn Yung, *Judged by the Company You Keep: An Empirical Study of the Ideologies of Judges on the United States Courts of Appeals*, 51 B.C. L. REV. 1133, 1136 (2010) ("[T]here has been a veritable explosion of empirical research about federal judges . . . ."); Michelle M. Mello & Kathryn Zeiler, *Empirical Health Law Scholarship: The State of the Field*, 96 GEO. L.J. 649, 651 (2008) ("In recent years, the field of empirical legal studies has grown exponentially."); Tracey E. George, *An Empirical Study of Empirical Legal Scholarship: The Top Law Schools*, 81 IND. L.J. 141, 142 (2006) (documenting the rise in empirical legal scholarship).

[27] Epstein & King, *supra* note 2, at 1.

law reviews during the eleven year period from 1990 to 2000.[28] In the subsequent period from 2001 to 2011, no fewer than 904 such papers were published in the same traditional printed law reviews, and many more were disseminated electronically.[29]

Professors Epstein and King were motivated to analyze empirical research in the legal community because of their observation that scholars were making unsupported inferences in their empirical research.[30] A decade later, that point is still worth discussing, but first it is worth stepping back and asking what is driving the explosion in empirical legal research. The supply forces are easy to see. We have experienced markedly lower costs in information collection, storage, and retrieval.[31] We have also experienced dramatically lower costs of computational software and computing power.[32] There are also more empirical researchers with better training in empirical research methodology.[33]

The increased demand for empirics is more subtle and difficult to document, but nevertheless something which would be commonly agreed on. By definition, empirical means working with observed data or experimental observations.[34] Observations and data are facts.[35] The inferences researchers make based on them might be flawed and not factual, but empirical research essentially involves collecting factual information

---

[28] *Id.* at 15–16.

[29] A search on Lexis-Nexis on August 8, 2012 for law review articles with "empirical" in the title published between 2001 and 2011 listed 904 publications. The same search for just the previous three years in the Social Science Research Network (SSRN) yielded 2,865 electronic abstracts and manuscripts.

[30] *See* Epstein & King, supra note 2, at 17 (stating that their comprehensive review of law review articles found that all of them had at least one mistake in their rules of inference without exception).

[31] *See* CARL SHAPIRO & HAL R. VARIAN, INFORMATION RULES: A STRATEGIC GUIDE TO THE NETWORK ECONOMY 20–21 (1999) ("[W]ith recent advances in information technology, the cost of distributing information is falling . . . . Information delivered over a network in digital form exhibits the first-copy problem in an extreme way: once the first copy of the information has been produced, additional copies cost essentially nothing.").

[32] *See, e.g.*, James Lindgren, *Predicting the Future of Empirical Legal Studies*, 86 B.U. L. REV. 1447, 1453 (2006) ("With today's computing power, most statistical analyses are almost instantaneous. In 1968, a single multiple regression analysis took all summer to do on a calculator. Now the same analysis would usually take less than a second to do on an ordinary laptop."); David Colander, *New Millennium Economics: How Did It Get This Way and What Way Is It?*, 14 J. ECON. PERSP. 121, 131 (2000) (observing that computing costs have dropped).

[33] *See* Carl E. Schneider & Lee E. Teitelbaum, *Life's Golden Tree: Empirical Scholarship and American Law*, 2006 UTAH L. REV. 53, 58 (2006) ("Law professors today do more empirical research than ever before. This is partly because there are many more scholars under much more pressure to do much more scholarship.").

[34] *See* Epstein & King, *supra* note 2, at 2 ("The word 'empirical' denotes evidence about the world based on observation or experience.").

[35] *See id.* at 3 ("[D]ata . . . is just a term for facts about the world.").

and using it to draw conclusions.[36] Forty years ago it might have been acceptable to make the argument that spending more money on teachers' salaries would improve the performance of school children without any supporting facts.[37] In the current environment, however, arguments that lack supporting data are easily dismissed.[38] We have come to expect and require data to support requests for resources.[39]

How does the state of empirical legal research compare with a decade ago? There has been much progress, but the legal community is still far behind researchers in the social and natural sciences.[40] Progress can be seen through the increase in law school faculty with Ph.D.s who have extensive training and experience in research methodology.[41] Progress can also be seen by the increase in law school course offerings and cross-listed courses covering statistical inference and quantitative methods.[42] However, these are still not commonplace.[43] Law school remains the only professional program which does not require statistics in the curriculum.[44] Researchers using empirical methods and quantitative models

---

[36] *Cf.* Klock, *Finding Random Coincidences*, *supra* note 10, at 1016 ("Statistical inference is the process of making inferences about a population based on a sample.").

[37] *See* Harold M. Baron, *Race and Status in School Spending: Chicago, 1961–1966*, 6 J. HUM. RESOURCES 3, 20 (1971).

[38] *See* Caroline M. Hoxby, *Does Competition Among Public Schools Benefit Students and Taxpayers?*, 90 AM. ECON. REV. 1209, 1236–37 (2000) (empirical research has shown that it is not increased spending that leads to higher achievement in students but rather the increase in school choices).

[39] *See, e.g.*, Spencer Overton, *Voter Identification*, 105 MICH. L. REV. 631, 631 (2007) (suggesting that empirical analysis of costs and benefits should be conducted before changing election laws).

[40] *See* Lee Epstein et al., *On the Effective Communication of the Results of Empirical Studies, Part II*, 60 VAND. L. REV. 801, 846 (2007) (observing that law professors are increasingly using data and performing quality work, but still are behind the social and statistical sciences in effective communication of empirical results).

[41] *Cf.* George, *supra* note 26, at 152 (creating a ranking of the top forty law schools based on the proportion of tenure-track faculty with a doctorate in a social science, a ranking that did not exist earlier, presumably due to the lack of a substantial number of law faculty with doctorates in social sciences).

[42] *See* Seth Freeman, *Bridging the Gaps: How Cross-Disciplinary Training with MBAs Can Improve Transactional Education, Prepare Students for Private Practice, and Enhance University Life*, 13 FORDHAM J. CORP. & FIN. L. 89, 102–03 (2008) (remarking on major commitments to cross-disciplinary training at NYU Law School, Stanford Law School, and University of Pennsylvania Law School).

[43] *See* Robert J. Rhee, *The Madoff Scandal, Market Regulatory Failure and the Business Education of Lawyers*, 35 J. CORP. L. 363, 388 (2009) (observing that although law students can take business classes at most institutions, they seldom do).

[44] *See* Steven B. Dow, *There's Madness in the Method: A Commentary on Law, Statistics, and the Nature of Legal Education*, 57 OKLA. L. REV. 579, 579 (2004) ("Professional legal education is unique among all of the university graduate-level programs in the natural and social sciences in not requiring at least a basic level of competency in statistics and quantitative methods."); Klock, *Finding Random Coincidences*, *supra* note 10, at 1063 (remarking that statistics is required in other professional schools such as business and education).

continue to make mistakes and the consumers of empirical research methods appear not to catch them.[45]

To take one example, a recent paper by William Landes and Richard Posner uses regression analysis to examine whether aspects of judicial behavior might be predictable based on observable variables.[46] Their dependent variable is the fraction of conservative votes in non-unanimous cases.[47] The variable is logically constrained to lie between zero and one, but using ordinary regression can yield predicted values which are negative or greater than one.[48] Hence an alternative methodology should have been used, or at least there should have been discussion about the potential flaws in the methodology.[49]

B. *Limits to Empiricism*

Although empirical analysis can be powerful and many of us encourage more of it, there are limits to empiricism.[50] Some scholars do not understand the limits and thus conduct their empirical research poorly.[51] Some scholars expect too much from data without a theory, or from theories that do not specify the precise form of the function that

45 *See, e.g.*, Mark Klock, *Contrasting the Art of Economic Science with Pseudo-Economic Nonsense: The Distinction Between Reasonable Assumptions and Ridiculous Assumptions*, 37 PEPP. L. REV. 153, 157–61 (2010) (taking one law professor to task for using a quantitative economic model to prove a desired result when the model contained inherently ridiculous and logically flawed assumptions) [hereinafter Klock, *Contrasting Economic Science*].

46 William M. Landes & Richard A. Posner, *Rational Judicial Behavior: A Statistical Study*, 1 J. LEGAL ANALYSIS 775, 775–76 (2009).

47 *Id.*; Table 3 at 782.

48 One way to understand the nature of the problem is to consider that ordinary regression often assumes normally distributed random variables underlying the model. *See* FINKELSTEIN & LEVIN, *supra* note 22, at 405 ("[I]t is frequently assumed that data are normally distributed . . . ."). Since the logical range of the dependent variable is bounded by the interval zero to one, a random variable with a normal distribution is impossible. *See also*, Jeff Yates & Elizabeth Coggins, *The Intersection of Judicial Attitudes and Litigant Selection Theories: Explaining U.S. Supreme Court Decision-Making*, 29 WASH. U. J.L. & POL'Y 263, 292 n.98 (2009) (acknowledging a potential problem using ordinary least squares when the dependent variable is a proportion bounded by zero and one).

49 An alternative methodology would be truncated regression. *See* CHRISTOPHER F. BAUM, AN INTRODUCTION TO MODERN ECONOMETRICS USING STATA 259 (2006) ("I now discuss a situation where the response variable is not binary or necessarily integer but has limited range. . . . Modeling LDVs [limited dependent variables] by OLS will be misleading.") (introducing discussion of truncated regression). It is certainly possible that this alternative methodology would not have produced different results, but without robustness tests or revelation of more details of the ordinary regression results we cannot ascertain this.

50 *See* Nancy C. Staudt & Tyler J. VanderWeele, *Methodological Advances and Empirical Legal Scholarship: A Note on Cox and Miles's Voting Rights Act Study*, 109 COLUM. L. REV. SIDEBAR 42, 46 (2009) ("[A]ll empirical researchers, *must* make assumptions about their data before estimating causal relationships and reporting empirical results.").

51 *See id.* ("While empirical researchers must, and always do, make assumptions about their data, these assumptions are almost always left unstated.").

relates the dependent variable to the explanatory variables.[52] For example, multiple regression analysis is one of the most commonly utilized empirical models.[53] Results of multiple regression have been introduced into evidence in countless trials.[54] A standard piece of rote memory from Ph.D. programs is that regression produces the "best linear unbiased estimates," and most doctoral students are taught the mathematical proof of this.[55] But the proof requires several assumptions, one of which is that the true model is known and is correctly specified.[56] So, if Y is in reality created by Model I, which is $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, and we instead estimate Model II, which is $Y = \alpha + \beta_1 X_1 + \varepsilon$, the regression estimates of β_1 will not be the best linear unbiased estimates.[57] Yet there are few, if any, situations in the social sciences where we can state with certainty that we know the true model *a priori*.[58] This causes other scholars to become disillusioned with empirical research and to devote time to criticizing it.[59]

Most educated people do not believe in alchemy.[60] Yet when it comes to empirical analysis and statistical models, many believe that something akin to alchemy is possible.[61] They believe that if a lot of data and numbers are fed into a computer, some black box alchemy-like process enables the computer to spew forth answers to important questions.[62] Will interest rates and gold prices rise or fall? Which potential

---

52 *See id.* ("[Empirical] authors . . . must presume that . . . some . . . set of relationships exists when specifying a model and interpreting the estimates from a regression . . . .").

53 *See* Richard Scheines, *Causation, Statistics, and the Law*, 16 J.L. & POL'Y 135, 159 (2007) ("When the measure of association used is correlation, then by far the most commonly used statistical technique for adjusting for confounders is multiple regression.").

54 *See, e.g.*, McClesky v. Kemp, 481 U.S. 279, 294 (1987) ("[T]his Court has accepted statistics in the form of multiple-regression analysis . . . .").

55 *See, e.g.*, Mark Thomas, *Material for Lecture I on Tuesday 1/4/11*, Economics 421 – Econometrics, http://economistsview.typepad.com/economics421/lectures/ (last visited Nov. 12, 2012) (covering the assumptions required for regression estimates to be best linear unbiased (BLU) and the Gauss-Markov Theorem).

56 *See* DAMODAR N. GUJARATI, BASIC ECONOMETRICS 66 (3d ed. 1995) (explaining that the model must be correctly specified).

57 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1057 (excluding relevant variables will result in biased estimates).

58 *See id.* at 1023 (suggesting that experts never know the true model with certainty).

59 *See, e.g.*, Tom Ginsburg, *Ways of Criticizing Public Choice: The Uses of Empiricism and Theory in Legal Scholarship*, 2002 U. ILL. L. REV. 1139, 1163 (2002) ("Legal scholarship is not primarily about empirical prediction. . . . [T]he key distinction of legal scholarship is its normative character. Legal scholarship is addressed to legal decision makers, with particular emphasis on judges who 'speak the same language' of the legal scholar.").

60 *See* David F. Hendry, *Econometrics—Alchemy or Science?*, 47 ECONOMICA 387, 387 (1980) (describing the pejorative connotations of alchemy).

61 *Cf.* Klock, *Finding Random Coincidences*, *supra* note 10, at 1008 ("[C]ommentators and reporters frequently give too much weight to statistics and treat them as actual facts rather than mere estimates which might not be valid or reliable for inferential reasoning.").

62 *See id.* at 1064 ("Statisticians are not alchemists and cannot create information out of thin air any more than they can create gold out of iron. They can feed numbers into a com-

jurors will vote to convict a criminal defendant or award large damages to a plaintiff? It should seem obvious that there is no computer or process that can provide accurate answers to such questions without inputting all of the information required to correctly answer them. There is an old adage in computer modeling that translates as inputting garbage produces output that is garbage.[63] Yet many people have the expectation that if we just put large enough amounts of garbage data into the computer, the computer will somehow miraculously produce high quality output.[64]

Professors Epstein and King refer to this belief in miracles as reification.[65] They observe that individuals reify numbers and treat them as something unalterable from a divine source, when the numbers are often merely rough approximations.[66] We might estimate the mean value of a distribution to be one, but if our 95% confidence interval around that estimate ranges from zero to two, then we do not have a very precise estimate of the mean.[67] Yet individuals will focus on the value of one as the correct value, ignoring the fact that it is no more than a crude approximation.[68]

The problem becomes much more complex when we attempt to condense a multi-dimensional concept into a one-dimensional measurement.[69] Intangible concepts, such as liberal and conservative, involve many dimensions, and efforts to create a simple ordinal measure of these concepts to rank Justices as more or less conservative inherently involve

---

puter and get numbers out, but the ingredients for valid information in the output must be in the input."); Edward E. Leamer, *Let's Take the Con Out of Econometrics*, 73 AM. ECON. REV. 31, 33 (1983) ("It would be a remarkable bootstrap if we could determine the extent of the misspecification from the data."); *id.* at 36 ("[D]ata alone cannot reveal the relationship . . . .").

63 *See* THOMAS H. WONNACOTT & RONALD J. WONNACOTT, REGRESSION: A SECOND COURSE IN STATISTICS 93 n.10 (1987) (describing the "garbage in, garbage out" approach).

64 *Cf.* Mark Klock, *Two Possible Answers to the Enron Experience: Will It Be Regulation of Fortune Tellers or Rebirth of Secondary Liability?*, 28 J. CORP. L. 69, 73 (2002) (explaining that the precise information people seek is often impossible to get because the calculations require estimates of future events, which cannot be confirmed in the present) [hereinafter Klock, *Two Possible Answers*].

65 *See* Epstein & King, *supra* note 2, at 28 & n.71 ("Reification is one of the oldest statistical mistakes on record.").

66 *Cf. id.* at 28 (describing how a group of empirical legal researchers created a profile of an average juror without any evidence that this profile represented "a majority of jurors, a few jurors, or any jurors at all").

67 *Cf.* Fischer Black, *The Trouble with Econometric Models*, FIN. ANALYSTS J., Mar.-Apr. 1982, at 29, 34 ("Even the best forecasts may not be very good.").

68 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1016 ("People have a strong tendency to treat statistics and estimates as if they are the actual values . . . .").

69 *See* Mark Klock, *Financial Options, Real Options, and Legal Options: Opting to Exploit Ourselves and What We Can Do About It*, 55 ALA. L. REV. 63, 109 (2003) ("[Measurement of] a multidimensional concept . . . cannot be implemented without creating an arbitrary and capricious scale.") [hereinafter Klock, *Financial Options*].

some arbitrary decisions.[70] If a Justice votes to strike down a law against protesting too close to an abortion clinic, is that a liberal vote for protecting the First Amendment or a conservative vote for empowering abortion protestors?[71] What if some conservative Justices vote to declare the law unconstitutional for the purpose of furthering a conservative cause of supporting abortion protestors, but use a liberal cause of free speech to justify the result?[72]

This limitation of quantitative measures can be seen more easily with more concrete examples, such as risk and size.[73] Risk has two dimensions: the probability of being different and the magnitude of the difference.[74] Imagine two different games. The first game pays you a dollar if a fair coin flip comes up tails and two dollars if it comes up heads, for an average payout of $1.50. The actual payout is always different from $1.50, but only by fifty cents. Another game pays you $1.49 99.9898% of the time and $100 0.0102% of the time. This game will nearly always pay the same amount, but has a very small chance of paying a lot more. It is not obvious which game is riskier since one has a high probability of a small difference while the other has a low probability of a large difference.[75] Measures of risk can be constructed that will make the first game seem riskier, and measures of risk can be constructed that will make the second game appear riskier, but the choice of risk measure is arbitrary.[76]

Size is commonly measured by weight, length, and volume.[77] For different purposes and different types of objects, one measure might seem superior to others.[78] However, if we are attempting to order different objects by size, choosing a specific dimension might create a strange

---

[70] *See* Klock, *Two Possible Answers*, *supra* note 64, at 95 (explaining that mapping multidimensional concepts into a one-dimensional measure is problematic because there is no unique mapping form).

[71] *Cf.* Hill v. Colorado, 530 U.S. 703, 714–16 (2000) (presenting the question of the constitutionality of a Colorado law restricting free speech in close proximity of health care facilities, including abortion clinics).

[72] *Cf. id.* at 741 (Scalia & Thomas, JJ., dissenting) (voting to declare restriction on speech near abortion clinics unconstitutional).

[73] *See* Klock, *Two Possible Answers*, *supra* note 64, at 95 (using risk and size to illustrate the difficulty of creating a consistent single ordering based on multiple attributes).

[74] *See id.*

[75] *Cf. id.* at 95–96 (giving a similar example where one distribution has a larger probability of being different from the average value, while the other has a larger probability of deviating from the average value by a greater amount).

[76] *Cf. id.* at 96 (explaining that the choice of which distribution is riskier is arbitrary).

[77] *See id.* at 95 ("Size can involve attributes such as length, width, thickness, mass, and volume.").

[78] For example: wrestlers are grouped by weight class; rope is measured by length; milk is sold by volume.

ordering.[79] One solution is to create a measure of size that combines weight, length, and volume.[80] This solution comes with its own problem though.[81] Once we use our new measure of size to describe objects, we lose all of the other information contained in the three original measures.[82] Furthermore, our method of combining the measures is arbitrary, and different methods of combining weight, length, and volume into a size measure will result in different orderings of the objects by size.[83]

Legal scholars have developed an appetite for quantitative empirical research.[84] Law generates a lot of data, and much of that data has not yet been subjected to quantitative analysis, which provides many interesting research opportunities.[85] Some scholars have advocated for a change in the traditional model of producing legal scholarship with a move towards peer-reviewed publications rather than student-edited publications.[86] There is a substantial body of literature criticizing and defending the law review model.[87] Professor Gregory Mitchell suggests a more practical solution.[88] He argues that the value added by the peer review process is

---

[79] *See* Klock, *Financial Options*, *supra* note 69, at 109 (giving a similar example regarding sorting policies by fairness). As stated there:

> [T]o evaluate whether a policy is more fair or less fair, we need to measure fairness. Two key principles of fairness are to treat equals equally and to treat unequals unequally. There is an immediately obvious tension between these principles when we recognize that people are similar and different in many dimensions and any classification system for individuals is necessarily arbitrary. Evaluating policies on fairness is like sorting heterogeneous objects from biggest to smallest without any clear purpose underlying the ordering.

Id.

[80] *See id.* ("We could arbitrarily sort based on weight, height, displacement, or any arbitrarily chosen function combining these aspects of size.").

[81] *See id.* ("The ordering will of course be dependent on arbitrary choices.").

[82] *See* Epstein & King, *supra* note 2, at 81 (describing the loss of information that necessarily occurs when creating a measure).

[83] *See* Leamer, *supra* note 62, at 38 (describing how inferential reasoning often rests on whimsical assumptions).

[84] *See* Susan Saab Fortney, *Taking Empirical Research Seriously*, 22 GEO. J. LEGAL ETHICS 1473, 1474–75 (2009) (commenting on the burgeoning field of empirical legal studies); Elizabeth Chambliss, *When Do Facts Persuade? Some Thoughts on the Market for "Empirical Legal Studies,"* 71 LAW & CONTEMP. PROBS., no. 2, 2008 at17, 22–23 (2008) (describing dramatic growth in empirical legal studies); George, *supra* note 26, at 141 (calling empirical legal scholarship the "next big thing in legal intellectual thought").

[85] *Cf.* Theodore Eisenberg, *Why Do Empirical Legal Scholarship?*, 41 SAN DIEGO L. REV. 1741, 1746 (2004) ("Across a broad range of legal issues, empirical studies can inform policymakers and the public. Legally trained social scientists have unique opportunities to enhance description and understanding of the legal system.").

[86] *See* Epstein & King, *supra* note 2, at 127–28.

[87] *See generally*, Cameron Stracher, *Reading, Writing, and Citing: In Praise of Law Reviews*, 52 N.Y.L. SCH. L. REV. 349 (2008) (discussing and refuting criticism of article selection by law reviews).

[88] *See* Mitchell, *supra* note 1, at 176 (suggesting "an alternative approach to improving empirical legal scholarship that may be more feasible than a move to peer review").

objectivity rather than validation.[89] He suggests that an easier way to obtain objectivity in the generation of empirical analysis is through strict disclosure requirements.[90] I follow this approach by disclosing the data and details of the methodology. I consider this investigation to be a model demonstrating how to collect data, analyze data, and report meaningful results.

II. MODELS OF VOTING BEHAVIOR

For reasons I will elaborate on later, without strong prior information, empirical analysis might not be capable of identifying the correct model of voting.[91] People often expect too much from data and statistical analysis.[92] Nevertheless, it is useful to have some models of voting behavior to provide a rough frame for the empirical analysis.[93]

This Article considers three distinct, but not mutually exclusive, models of voting: independent, cooperative, and vindictive. Economists usually assume that individual economic agents, such as households or voters, have a set of preferences that are independent of each other.[94] On some level, it might seem that this assumption is clearly false.[95] People

[89] *See id.* at 175 (arguing "that the primary benefit of peer review lies in its objectivity-forcing function: peer review compels the disclosure of important information about empirical research using a common methodological language so that the research may be subjected to critical scrutiny.").

[90] *See id.* at 176 ("[L]aw reviews can force objectivity into empirical legal scholarship by adopting a set of stringent disclosure requirements for reports of original empirical research, including disclosure of detailed information about methodology, data analysis, and the availability of raw data for replication and review.").

[91] *See* Leamer, *supra* note 62, at 36 ("[D]ata alone cannot reveal the relationship . . . . [W]e must resort to subjective prior information.").

[92] *See* Klock, *Two Possible Answers*, *supra* note 64, at 94 (suggesting that people expect statistical estimates to reveal the truth even though they cannot).

[93] *See* Mark Klock, *Are Wastefulness and Flamboyance Really Virtues? Use and Abuse of Economic Analysis*, 71 U. CIN. L. REV. 181, 195 n.88 (2002) (remarking that "[f]acts alone cannot explain events," and a theory or model helps explain a set of facts) [hereinafter Klock, *Wastefulness*].

[94] *See, e.g.*, Michael C. Jensen, *The Nature of Man*, *in* THE NEW CORPORATE FINANCE: WHERE THEORY MEETS PRACTICE 4, 5 (Donald H. Chew, Jr. ed., 2d ed. 1999) (stating that two assumptions commonly employed in the model of utility maximizing households are preference for more over less and independence of preferences across households).

[95] *See* Herbert Hovenkamp, *Arrow's Theorem: Ordinalism and Republican Government*, 75 IOWA L. REV. 949, 954–55 (1990). In criticizing the economic model of individual preferences, Professor Hovenkamp writes that:

> It treats legislators something like children selecting a single flavor of ice cream to be shared by all. Each child's preferences are strictly individual, and there is generally no reason to prefer the preferences of one child over those of another. Likewise, the children do not take the strengths of one another's preferences into account.

Id. at 955.

give money to charities and they share with relatives.[96] Some economists and biologists, however, believe that such giving is still motivated by self-interest.[97] There is selfishness when one gives to influence the perceptions of others about one's self; when one gives to feel good; and when evolutionary forces might induce people to give to relatives and communities to improve the chances that one's own genes, or genes very similar to one's own, survive.[98] Whether this assumption is correct or not, however, is not as important as whether it is a reasonable first approximation.[99] All models distort reality to some degree.[100] The art of good model building is to use assumptions that simplify some of the less important complexities of reality in order to highlight other relationships without grossly distorting those relationships.[101] The standard economic assumption of independence of preferences makes the mathematical analysis of such models more tractable.[102]

Some lawyers might be quick to suggest that independence of voting in the context of the Supreme Court is ridiculous since the Justices discuss the cases together and vote with the junior associate Justice voting first and the Chief Justice voting last.[103] If voting preferences are independent, why bother to discuss the case and attempt to persuade others, and why attach so much importance to the order of voting? On the other hand, such discussion may merely be a mechanism by which independent voters form their own decisions, as if they are talking out loud to themselves and considering and weighing all of the issues; just as registered voters listen to campaign debates, speeches, and commercials

---

96 *See* STEVEN D. LEVITT & STEPHEN J. DUBNER, SUPERFREAKONOMICS: GLOBAL COOLING, PATRIOTIC PROSTITUTES, AND WHY SUICIDE BOMBERS SHOULD BUY LIFE INSURANCE 104 (2009) ("We all witness acts of altruism, large and small, just about every day.").

97 *See* Arthur J. Robson, *The Biological Basis of Economic Behavior*, 39 J. ECON. LITERATURE 11, 23 (2001) ("One direct implication of biology that many economists would accept is altruism among close relatives." In other words, people share with close relatives because it is in their genetic interest to do so).

98 *See* LEVITT & DUBNER, *supra* note 96, at 124 ("Most giving is, as economists call it, *impure altruism* or *warm-glow altruism*. You give not only because you want to help but because it makes you look good, or feel good, or perhaps feel less bad.").

99 *See* MILTON FRIEDMAN, ESSAYS IN POSITIVE ECONOMICS 14–16 (1953) (arguing that the truth of a theory is unimportant if the theory accurately predicts reality).

100 *See* Kenneth G. Dau-Schmidt, *Economics and Sociology: The Prospects for an Interdisciplinary Discourse on Law*, 1997 WIS. L. REV. 389, 397 (1997) ("Every model or analysis of a problem is necessarily an abstraction from reality, ignoring some complication of life to focus on others.").

101 *See* Klock, *Contrasting Economic Science*, *supra* note 45, at 198 ("The art of good model-building lies in the ability to assume well.").

102 *See* Klock, *Wastefulness*, *supra* note 93, at 240 ("[Selfishness] is merely a simplifying assumption that produces tractable models with highly accurate predictions in many cases.").

103 *See* Tom C. Clark, *Internal Operation of the United States Supreme Court*, 43 J. AM. JUDICATURE SOC'Y 45, 50 (1959) ("After discussion of a case, a vote is taken. . . . [T]he formal vote begins with the junior Justice and moves up through the ranks of seniority, the Chief Justice voting last.").

and then reach their own decision;[104] and just as consumers sort through all sorts of marketing material before making purchases.[105] Indeed, many would be comforted by the idea that the world really works this way, with each Justice sincerely applying his best interpretation of the law to reach a non-political result and then aggregating across the results using a "majority rules" procedure.[106]

An interesting paradox results using rational, independent preferences in a democratic process.[107] A Nobel economist, Kenneth Arrow, mathematically proved that no system of aggregating rational, independent preferences, other than a perfect dictatorship, will guarantee that the aggregated preferences will also be rational.[108] The proof of this is known as Arrow's Impossibility Theorem because it proves the impossibility of constructing a democratic system of voting that will be consistently rational.[109] The proof is complex, but a simple example illustrates the idea. Rationality requires that if A is preferred to B, and B is preferred to C, then A must also be preferred to C.[110] That is, if a consumer prefers blue cars over red cars, and red cars over green cars, then the rational consumer must prefer a blue car over a green car.[111] With nine Justices, however, it is possible to have three prefer blue to red to green, three prefer red to green to blue, and the final three prefer green to blue to red. In such a scenario, if red is selected, two-thirds of the Justices would prefer blue. If blue is selected, two-thirds of the Justices would prefer green. And if green is selected, two-thirds would prefer red. This result can explain apparent instability in many political decisions.

---

[104] *Cf.* Citizens United v. FEC, 130 S. Ct. 876, 898 (2010) ("The right of citizens to inquire, to hear, to speak, and to use information to reach consensus is a precondition to enlightened self-government and a necessary means to protect it.").

[105] *See, e.g.*, Gary S. Becker & Kevin M. Murphy, *A Simple Theory of Advertising as a Good or Bad*, 108 Q. J. ECON. 941, 955 (1993) (providing an example of advertising that attempts to convince consumers that one brand of chicken is more valuable than another brand of chicken).

[106] *See, e.g.*, Jack M. Balkin, Bush v. Gore *and the Boundary Between Law and Politics*, 110 YALE L.J. 1407, 1409 (2001) (suggesting that the appearance of partisanship in Supreme Court decisions makes the Court's output unsavory).

[107] *See* HAL R. VARIAN, INTERMEDIATE MICROECONOMICS 634 (8th ed. 2010) ("[The] very plausible and desirable features of a social decision mechanism are inconsistent with democracy . . . .").

[108] *See id.*

[109] *See* Klock, *Bush v. Gore*, *supra* note 21, at 15 ("Professor Arrow's modern contribution is formally proving under very general conditions that it is *impossible* to create *any* democratic voting scheme that will result in rational social preferences.").

[110] *See* VARIAN, *supra* note 107, at 35–36 (explaining the transitivity axiom of consumer preferences and why it is reasonable).

[111] *See id.* at 36 (describing the peculiarity resulting from intransitive preferences when comparing three choices).

Professor Herbert Hovenkamp criticizes the independent preferences assumption of Arrow's Theorem.[112] Professor Hovenkamp argues that people are cooperative and will sacrifice their own weak preferences in order to yield to the strong preferences of others.[113] For example, if I have a slight preference for vanilla ice cream over chocolate, and my neighbor has a strong preference for chocolate over vanilla, then when we get together to freeze some homemade ice cream and only have one machine between the two of us (such that making both kinds is not an option), I will agree to chocolate rather than vanilla. In Hovenkamp's model of cooperative voting, aggregated preference orderings are much less likely to be intransitive than under the independent model because of cooperation and the willingness to yield when preferences are slight.[114]

One problem with an application of Hovenkamp's model to the Supreme Court is that many of the divisive, controversial, and important cases that come before the Court involve issues for which preferences are exceedingly strong, passionate, and uncompromising. Consider capital punishment for murderers, abortion on demand, waterboarding of terrorists, prayer in public schools, and similar issues. People tend to either find these positions acceptable or unacceptable, without much room for compromise. Many Justices, just like many voters, have strong views on these issues and the application of the Constitution to these issues.[115] These are not cases where it would be reasonable to expect people to compromise or yield their own judgments.

An alternative to cooperation is vindictiveness.[116] A group that feels passionate about one issue and has relatively weak preferences on other issues might deliberately vote against others in retaliation if they do not get the support they want on their major issue.[117] This is a well-known phenomenon in politics.[118] Assuming a diversity of preferences

---

112 *See id.* at 35 ("[I]f the consumer thinks that X is at least as good as Y and that Y is at least as good as Z, then the consumer thinks that X is at least as good as Z.").

113 *See* Hovenkamp, *supra* note 95, at 952 (questioning the reasonableness of independent preferences).

114 *See id.* at 952–53 (providing a hypothetical example of cooperative voting).

115 *See, e.g.*, Clarence Thomas, *Judging*, 45 U. KAN. L. REV. 1, 7 (1996) ("We as a nation adopted a written Constitution precisely because it has a fixed meaning that does not change.").

116 *See* Klock, *Contrasting Economic Science*, *supra* note 45, at 192–93 (stating that coalitions of voters can be vindictive when fighting for their cause).

117 *See, e.g.*, Anthony Lewis, *Friendly Persuasion*, N.Y. TIMES, Apr. 28, 2002, at F8 (describing how Lyndon Johnson "took disagreement personally"); David Nakamura, *Resignation's Reverberations: Thornton's Move Creates Local, State Intrigue*, WASH. POST, Oct. 13, 1999, at M07 (suggesting that, in certain instances, local politicians hold grudges and retaliate on other issues when given the chance).

118 *See* Klock, *Contrasting Economic Science*, *supra* note 45, at 193 ("Highly charged issues . . . serve as emotional battlefields where people hold strong and uncompromising beliefs, and thus are willing to vote against other groups' issues in retaliation for those groups' lack of support for their own.").

among individuals, and assuming a diversity of passions among individuals, the vindictive model of voting is likely to produce even more intransitive preference orderings than the independent voting model.[119]

Empirically testing these models of voting is problematic. If we make some arbitrary assumptions about the structures of the cooperative and vindictive models and of the SRCE Justices, we could design a test statistic that could inform us about whether the data is consistent with the models or highly unlikely to have been produced by them.[120] Without imposing more structure on the cooperative and vindictive models we cannot reliably test them.[121] Additionally, any structure that might be assumed is essentially conjured out of thin air.[122] Likewise, it is difficult to develop a definitive test of the independent voting model. If there is no correlation between the Justices' votes, it would be consistent with the independent voting model.[123] However, not surprisingly, the votes are not uncorrelated.[124] Just because the votes are not uncorrelated does not mean that the Justices do not vote independently, for another variable could be affecting the voting behavior of the Justices.[125] Using only the voting data, it is not possible to distinguish between the hypothesis that Rehnquist, Scalia, and Thomas secretly agree to vote as a block 90% of the time and the hypothesis that some omitted variable, such as conservative values, drives them towards the same result 90% of the time.[126]

Nevertheless, just because our models lack detailed structure does not mean we should abandon them.[127] The models still give us some

---

[119] *See id.* at 192–93 (explaining that voters with lexicographic preferences are not willing to compromise or trade their principles).

[120] *Cf.* Leamer, *supra* note 62, at 43 ("In order to draw inferences from data as described by econometric texts, it is necessary to make whimsical assumptions.").

[121] *See id.* at 36 ("A model with an infinite number of parameters will allow inference from a finite data set only if there is some prior information that effectively constrains the ranges of the parameters.").

[122] *See id.* at 37 (characterizing statistical inferences as opinions due to the whimsical nature of the assumptions on which they rest).

[123] *Cf.* FINKELSTEIN & LEVIN, *supra* note 22, at 31 (stating that in general, if two variables are independent then their correlation is zero).

[124] *See infra* Table 3.

[125] *See* Landes & Posner, *supra* note 46, at 787 (describing evidence that ideology matters in the Justices' votes).

[126] *Cf.* Klock, *Wastefulness*, *supra* note 93, at 195 n.88 ("A theory is a set of explanations which can be refuted or supported by facts, but cannot be proven to be true due to the impossibility of ruling out alternative explanations of the same facts. Logically, one theory cannot disprove another theory.").

[127] *See* David F. Hendry, ECONOMETRICS: ALCHEMY OR SCIENCE? ESSAYS IN ECONOMETRIC METHODOLOGY 1 (1993) ("Although important technical difficulties about the properties of tests and of model selection procedures based on sequential testing await resolution, model evaluation is a legitimate activity . . . .").

insight as to what we should be looking for.[128] Additionally, notwithstanding the gloomy warning about the limits of empirical analysis, we should not give up on empirics either.[129] There is no mathematical tool that can create information to answer questions in the absence of the required information,[130] but sifting through the data can provide insight even if it does not provide definitive answers.[131] Analysis of the data can give us quantitative measures of simple attributes that are indisputable.[132] These measures are more persuasive than qualitative statements.[133] The empirical analysis of the SRCE data provides simple quantitative measurements of voting behavior, and applies some complex modeling that is at least suggestive of underlying relationships in the data.

Scholars working in the area of empirical legal studies are not the only researchers facing difficult challenges.[134] When modern computing power became a reality, the field of empirical econometrics expanded.[135] There were expectations that one day large statistical models would be able to accurately predict the future of the economy.[136] Certainly there

---

128 *Cf.* Landes & Posner, *supra* note 46, at 779 ("We do not propose a formal economic model of judicial behavior, but in the next part we sketch an informal such model to guide our empirical analysis.").

129 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1064 (stating that although statistical analysis of data cannot identify the true model, statistics can still be used to effectively present and communicate information).

130 *See* Leamer, *supra* note 62, at 37 ("Because both the sampling distribution and the prior distribution are actually *opinions* and not *facts*, a statistical inference is and must forever remain an *opinion*.").

131 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1060 ("Where no well-conceived theories exist, the kind of quantitative analysis conducted is a useful exercise to investigate the stylized facts for use in building theoretical models to be tested with independent data.").

132 *See, e.g.*, Samuel R. Gross & Kent D. Syverud, *Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial*, 90 MICH. L. REV. 319, 380 (1991) (analyzing data and finding that "[i]f plaintiffs rather than their attorneys are required to advance trial costs (including attorneys' fees), and to bear the risk of failing to recover those costs, the trial rate will decline and the plaintiffs' success rate at trial will increase.").

133 For example, stating that Justice Thomas' voting record has a correlation with conservative dispositions of 0.575 while Justice Kennedy's voting record has a correlation with conservative dispositions of 0.265 is more informative than a qualitative statement that Justice Thomas is substantially more conservative than Justice Kennedy.

134 *Cf.* Epstein & King, *supra* note 2, at 17–18 ("In writing this, we do not mean to suggest that empirical research appearing in law reviews is always, or even usually, worse than articles in the journals of other scholarly disciplines.").

135 *See* Michael C. Lovell, *Data Mining*, 65 REV. ECON. & STAT. 1, 1 (1983) ("The efficiency with which data miners go about their work has increased considerably as a result of technological advance[s].").

136 *See* Herman O. Wold, *Econometrics as Pioneering in Nonexperimental Model Building*, 37 ECONOMETRICA 369, 369 (1969) ("Econometrics is seen as a vehicle for fundamental innovations in scientific method, above all in the development of operative forecasting procedures in nonexperimental situations.").

was much disappointment when these expectations went unfulfilled.[137] Yet even though people have learned that future economic conditions cannot be consistently predicted accurately, massive resources from both the public and private sectors continue to be invested into predicting future economic conditions.[138]

Some of the best clues to predict future economic conditions come from simplistic quantitative measures, such as changes in inventories.[139] The field of empirical legal studies can learn lessons from other empirical subjects.[140] One lesson is not to expect too much from the data.[141] Another lesson is not to get frustrated and give up.[142]

III. DATA AND METHODOLOGY

A. *Choice of Sample Period*

Selecting the SRCE should be an obvious choice for analysis. In order to have any chance of successfully learning about the effects of one variable on another in a complex system, it is necessary to isolate the effects of other variables by holding them constant.[143] So, if we want to uncover the effects of Justice Kennedy's persuasive power on Justice Thomas, we need to hold the composition of the Court constant. From Justice Breyer's assumption of office on August 3, 1994, until the death of Justice Rehnquist on September 3, 2005, there was no change in the composition of the Court.[144] This was the second longest time period in

---

137 *See, e.g.*, Hendry, *supra* note 60, at 402 ("It is difficult to provide a convincing case for the defence [sic] against Keynes' accusation almost 40 years ago that econometrics is *statistical alchemy* since many of his criticisms remain apposite.").

138 *See id.* at 389 ("Substantial resources have been devoted to empirical macroeconometric models which comprise hundreds or even thousands of statistically calibrated equations, each purporting to represent some autonomous facet of the behaviour [sic] of economic agents such as consumers and producers, the whole intended to describe accurately the overall evolution of the economy.").

139 *Cf.* BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM, 97TH ANNUAL REPORT 13 (2010) (describing inventory investment).

140 *See, e.g.*, Epstein et al., *supra* note 40, at 846 (suggesting that law professors working on empirical legal scholarship adopt methods used in the literature of social and statistical sciences).

141 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1009 ("Classical statistical theory begins with the premise that one knows the true model independent of the data . . . .").

142 *See* Hendry, *supra* note 60, at 396 ("That the subject is exceedingly complicated does not entail that it is hopeless.").

143 *See* Mark Klock, *Dead Hands–Poison Catalyst or Strength-Enhancing Megavitamin? An Analysis of the Benefits of Managerial Protection and the Detriments of Judicial Interference*, 2001 COLUM. BUS. L. REV. 67, 79 (2001) (explaining the role of *ceteris paribus* in economic models).

144 Michael Allan Wolf, *Supreme Guidance for Wet Growth: Lessons from the High Court on the Powers and Responsibilities of Local Governments*, 9 CHAP. L. REV. 233, 234 (2006).

history with no change on the Court,[145] the longest being the period between 1812 and 1823.[146] Of course the Court took fewer cases in those days and there were fewer Justices,[147] so the SRCE provides the richest source of voting data holding the composition of the Court constant. My sample contains voting data on 920 published opinions by the United States Supreme Court.

One of the most basic rules of empirical research is that if the researcher wants to infer something about the relationship between X and Y, all other variables must be held constant, or controlled in a way that allows the relationship between X and Y to be revealed.[148] Suppose that a farmer observes the following data on crop yield and rainfall for eight years:[149]

Yield (bushels per acre)	Total Spring Rainfall (inches)
60	8
50	10
70	11
70	10
80	9
50	9
60	12
40	11

Given this data, the farmer might infer that more rainfall resulted in a smaller crop yield because a regression of yield on rainfall results in an estimate that an additional inch of rain lowers yield by 1.67 bushels.[150]

---

145 *See Nine Justices, Ten Years: A Statistical Retrospective*, 118 HARV. L. REV. 510, 510 & n.1 (2004) (observing that "[a] seven-Justice Court sat together for more than a decade only once, from Justice Joseph Story's appointment in 1812 [term starting in 1811] to Justice Brockholst Livingston's death in 1823.").

146 *Id.* at 510 n.1.

147 *See id.* (indicating that only seven Justices sat on the Court during Justice Story's tenure as Junior Associate Justice). In the days of the Marshall Court, the Court met only during February and March because of the light work load. A Lexis-Nexis search on Court opinions issued in 1812 identified thirty-two published opinions all dated between February 25 and March 14.

148 *See* WONNACOTT & WONNACOTT, *supra* note 63, at 8 (explaining that to study a relationship between two variables one needs to hold all other variables constant, and where that cannot be done, one needs to control for the other variables by compensating so as to obtain the same answer as if the other variables were held constant).

149 This example is taken from *id.* at 99.

150 *Id.*

Now suppose we add an additional variable and reveal that the farmer's initial crude analysis forgot to control for temperature:[151]

Yield (bushels per acre)	Total Spring Rainfall (inches)	Average Spring Temperature, °F
60	8	56
50	10	47
70	11	53
70	10	53
80	9	56
50	9	47
60	12	44
40	11	44

With this additional information we can see that rainfall does increase crop yield by an average of 5.71 bushels per inch, and that the first anomalous inference was made because crop yield is also positively affected by temperature, and that the complete data reveals that large amounts of rainfall occurred in the colder years.[152]

In addition to controlling for all variables, the model must remain consistent during the entire period to ensure valid statistical inferences.[153] There cannot be any structural change.[154] In the farming example, if there had been a breakthrough discovery in a revolutionary new type of fertilizer in the middle of the study period, the inferences drawn from the data would be flawed.[155] A change in the composition of the Supreme Court is an example of such a structural change.[156] The relationship between Justice Scalia and Justice Thomas could be different depending on whether Justice Blackmun or Justice Kagan is on the bench. If we want to examine that relationship, it is essential to control for the composition of the remainder of the Court. Failure to recognize

---

151 *See id.*

152 *See id.* at 99–100.

153 *See* WILLIAM H. GREENE, ECONOMETRIC ANALYSIS 130 (5th ed. 2003) ("In specifying a regression model, we assume that its assumptions apply to all the observations in our sample.").

154 *See id.* (explaining that if a structural change occurs, the same regression model will not apply for all observations).

155 *Cf. id.* (using an example of structural change in the gasoline market stemming from large oil price shocks).

156 *See* Daniel E. Ho & Kevin M. Quinn, *Did a Switch in Time Save Nine?*, 2 J. LEGAL ANALYSIS 69, 72 (2010) (using statistics to show that a structural shift in the Court occurred after Franklin Roosevelt's appointments to the Court).

the danger of a structural change in the Court's composition can lead to flawed results and thus lead to improper inferences from those results.[157]

One might argue that a particular Justice will not be influenced by other Justices on the Court—this is essentially the independent voting model.[158] Such an argument is too simplistic. Justices only vote on the cases they select to hear.[159] A change in the composition of the Court could affect the cases the Court selects.[160] Thus, even if each Justice votes independently of the other Justices, a change in the composition of the Court nevertheless creates an important structural change.[161] We are therefore safest in only drawing inferences based on a stable Court. Since the SRCE provides us with the largest amount of data on a stable Court, it is the best place to conduct an empirical analysis of voting.

B. Data Collection

Data was collected for all opinions involving the Supreme Court's cases of original jurisdiction, cases brought on appeal, and cases granted a writ of certiorari. Opinions regarding denial of certiorari, motions for reconsideration, applications for stays, applications to vacate stays, etc. were disregarded. Cases that were disposed of when the court dismissed a writ of certiorari as improvidently granted were also disregarded.

For each case I recorded the date, citation, and case number. The case number is simply collected as a redundant method of identifying the opinion in case of an error in recording the citation. Part of the appeal of empirical research is the transparency of the data collection and methodology, both of which subject the investigator to the scrutiny of other researchers attempting to replicate and confirm or dispute the analysis.[162] I also recorded the vote of each of the nine Justices. The votes were recorded as follows: "1" if the Justice voted with the majority, "0" if the

---

[157] *See, e.g.*, Landes & Posner, *supra* note 46, at 781 (analyzing a sample of Supreme Court Justice voting that includes forty-three different Justices' votes between 1937 and 2006).

[158] *Cf. id.* at 789 ("Supreme Court Justices do not acknowledge that any of their decisions are influenced by ideology rather than by neutral legal analysis.").

[159] *See* David C. Thompson & Melanie F. Wachtell, *An Empirical Analysis of Supreme Court Certiorari Petition Procedures: The Call for Response and the Call for the Views of the Solicitor General*, 16 GEO. MASON L. REV. 237, 241 (2009) ("Of the 8,517 petitions filed in the Court's 2005–06 Term . . . only 78 were granted argument (0.9%).").

[160] *See* Michael S. Greve & Jonathan Klick, *Preemption in the Rehnquist Court: A Preliminary Empirical Assessment*, 14 SUP. CT. ECON. REV. 43, 60–63 (2006) (documenting a large shift in the pattern of granting certiorari in the Rehnquist Court before and after the appointment of Justice Breyer in 1994).

[161] *See id.* This follows from the fact that a change in the composition of the Court changes the case selection process, which changes the observation generating process.

[162] *See* Mitchell, *supra* note 1, at 176–77 ("[D]isclosure norms would make empirical legal research more amenable to intersubjective review and testing and would go far toward making this body of research a more objective, respected, and productive form of scientific dialogue."). Upon request I will provide copies of my data file to other academic researchers.

Justice dissented, and "-1" if a Justice did not participate. Of course there are complications when a Justice concurs in part and dissents in part. In such situations I read the opinion and decided whether the vote should count as a dissent or not. In most instances such opinions involved a strong support for a different disposition of some aspect of the case and were treated as dissents, but if the partial dissent involved a minor procedural matter it was treated as concurring with the majority. There are also a number of cases (less than one percent) in which a majority of the Justices dissented in part. In these cases, I classified the partial dissents that most closely aligned with the plurality as part of the majority in order to avoid the anomalous result of having a majority of Justices dissenting.

In addition to this data, I also collected three more indicator variables regarding the cases. One variable indicates whether the case is a criminal matter or not. This criminal variable is recorded as a "1" if the case was criminal in nature and "0" otherwise. Cases involving deportation proceedings based on an underlying crime are treated as criminal, as well as disputes over sentencing, parole, solitary confinement, etc. Another variable indicates whether the Court affirmed the lower court. This variable is recorded as a "1" if the lower decision is affirmed and "0" otherwise. A "0" does not necessarily mean the Court reversed the lower court because a decision to vacate and remand would also be recorded as a "0". The more problematic decisions are the ones where the Court affirms in part and reverses in part. Again, these decisions require a close read and evaluation to treat the judgment as affirmed or not. Normally, a partial reversal would be recorded as a "0," meaning not affirmed.

Finally, the third and most problematic variable indicates whether the disposition of the case is conservative (recorded as "1") or not (recorded as "0"). This creates problems on multiple levels. As a researcher, I try to remain detached from the pros and cons of conservative and liberal views, yet I must disclose the basis of the methodology for classifying the dispositions of the cases.[163] Although this is likely to offend some people, perhaps all, I characterize conservatives as cold-hearted towards the plaintiffs in a wrongful death action and liberals as loving criminals. I apologize for these rough and unflattering characterizations and note that many conservative Justices vote for liberal dispositions and many liberal Justices vote for conservative dispositions.[164]

[163] *Cf.* Epstein & King, *supra* note 2, at 9 ("An attorney who treats a client like a hypothesis would be disbarred; a Ph.D. who advocates a hypothesis like a client would be ignored.").

[164] Indeed, in my sample the dispositions of the unanimous decisions were coded conservative (liberal) at a rate of 59% (41%). *See infra* Table 1.

Another problem is that some cases are difficult to classify as conservative or liberal.[165] Patent disputes are a common example, where decisions do not appear to correlate with whether a Justice is liberal or conservative.[166] Border disputes or water rights disputes between states are other examples.[167] Nevertheless, a decision needs to be made as to whether the disposition is more liberal or conservative. Fortunately, many of these difficult to classify cases are 9-0 decisions and will not play much role in analyzing how conservative or liberal outcomes affect marginal cases.[168] Perhaps it should not be surprising that unanimity is more common in cases that do not have strong political undertones.[169]

For criminal cases, the determination of whether a disposition is conservative or liberal is fairly straightforward. Decisions favoring the prosecution are conservative, while decisions that favor the defense are liberal. There are three types of exceptions to this general rule. First, if the crime is merely possession of a handgun and the statute is found to be unconstitutional, the disposition is coded as conservative because gun rights are considered a conservative value.[170] Second, if the crime involves burning crosses and the conviction is upheld, this is recorded as a liberal disposition even though it goes against the criminal because I deem liberal civil rights values to trump other liberal values.[171] Third, if the crime involves securities fraud and the Court reverses the circuit court's reversal of a conviction, the decision is recorded as a liberal outcome since an expansive construction of the federal securities laws is associated with liberal values.[172]

---

165 *See* Landes & Posner, *supra* note 46, at 777 (noting difficulty in coding decisions as liberal or conservative).

166 *See generally* Holmes Grp., Inc. v. Vornado Air Circulation Sys., Inc. 535 U.S. 826 (2002) (case filed in federal district court resulting in a counter-claim involving patent infringement with an appeal filed in the Federal Circuit). The Court vacated the circuit court's decision and I coded this disposition as conservative because the decision was based on a determination that the Federal Circuit lacked jurisdiction. *See id.* at 834.

167 *See generally* Arizona v. California, 530 U.S. 392 (2000).

168 *See, e.g.*, *Holmes*, 535 U.S. at 834 (Justices voting unanimously for the disposition, but with two concurring opinions).

169 *Cf.* Lee Epstein et al., *Dynamic Agenda-Setting on the United States Supreme Court: An Empirical Assessment*, 39 HARV. J. ON LEGIS. 395, 409–10 (2002) (showing that unanimous decisions are much less likely to be scrutinized by Congress).

170 *See e.g.*, United States v. Lopez, 514 U.S. 549, 550, 567 (1995) (majority decision by Justice Rehnquist, joined by Justices O'Connor, Scalia, Kennedy, and Thomas).

171 *See e.g.*, Virginia v. Black, 538 U.S. 343 (2003) (affirming the reversal of a defendant's conviction for cross-burning).

172 *See e.g.*, United States v. O'Hagan, 521 U.S. 642 (1997) (holding a defendant criminally liable for insider trading). *See generally* Mark Klock, *What Will It Take to Label Participation in a Deceptive Scheme to Defraud Buyers of Securities a Violation of Section 10(b)? The Disastrous Result and Reasoning of* Stoneridge, 58 KAN. L. REV. 309 (2010) (providing a discussion suggesting that conservative Justices favor a narrow construction of the securities laws and liberal Justices favor a broad remedial interpretation).

The rules for classifying the disposition of non-criminal cases are more complex and sometimes more discretion is required because the cases can have multiple dimensions. Dispositions broadening free speech rights are liberal, and those restricting speech are marked as conservative. Dispositions favoring employees, unions, the disabled, Native American Indian tribes, class-action plaintiffs, and debtors are marked as liberal. Dispositions favoring employers, businesses, mandatory arbitration agreements, private property rights, and creditors are marked as conservative. Dispositions that are reached not on the merits but by arguments involving lack of standing or lack of jurisdiction tend to be recorded as conservative outcomes, whereas dispositions that involve an expansive interpretation of federal jurisdiction tend to be recorded as liberal outcomes.

C. Remarks about Statistical Methodology

The difficulty of classifying dispositions is a good example of the limitations of empirical work.[173] Empirical work usually entails a model, but models are necessarily abstractions of reality.[174] Good models simplify reality in a way that allows us to better understand relationships of interest.[175] Consider the economist's model of demand as one example. Demand for a good is inversely related to the price of the good.[176] Since the price of a good is what one has to give up to obtain it, the less one has to sacrifice to get the good, the more desirable the good will be, and the more it will be demanded.[177] The result is not an as-

---

[173] *See* Epstein & King, *supra* note 2, at 85–86 (discussing the difficulty of classifying case dispositions).

[174] Consider the following excerpt from a popular text:

> Because all models simplify reality by stripping part of it away, they are abstractions. Critics of economics often point to abstraction as a weakness. Most economists, however, see abstraction as a real strength.
>
> . . . .
>
> Like maps, economic models are abstractions that strip away detail to expose only those aspects of behavior that are important to the question being asked. . . .
>
> But be careful. Although abstraction is a powerful tool for exposing and analyzing specific aspects of behavior, it is possible to oversimplify. . . .
>
> The key here is that the appropriate amount of simplification and abstraction depends upon the use to which the model will be put. To return to the map example: You don't want to walk around San Francisco with a map made for drivers—there are too many very steep hills!

KARL E. CASE & RAY C. FAIR, PRINCIPLES OF MICROECONOMICS 10–11 (5th ed. 1999).

[175] *See id.*

[176] *See* Mark Klock, *Unconscionability and Price Discrimination*, 69 TENN. L. REV. 317, 320 (2002) ("As a general rule, as the price of an item falls, the amount that people are willing to buy and consume—that is, the quantity demanded—increases.").

[177] *See* COOTER & ULEN, *supra* note 25, at 25 ("This result is the famous *law of demand*.").

sumption itself but is a derivation from a few simple axioms.[178] The three axioms are the following: consumers have a complete set of preference orderings; the preference orderings are transitive; and more is preferred to less.[179] There could be exceptions that probe the rule, but these axioms have held up remarkably well in experiments involving normal people, cognitively impaired people, and laboratory rats and pigeons.[180]

The model of demand helps us to better understand the role that price plays in affecting purchases of a good, but it abstracts from reality because it ignores (or assumes constant values for) all other important variables, such as income, prices of other goods, tastes, fashion, etc.[181] A change in any of these other variables will change the relationship between demand and price.[182] Since these variables change constantly, the relationship between demand and price is not stable, but the model still shows that price is an important determinant which affects the quantity demanded in a negative way.[183] While holding everything else constant, as the price of a good rises, people tend to substitute different goods into their consumption bundles.[184] Some goods are more easily substituted than others.[185] Rising beef prices will lead to more consumption of poultry, pork, fish, and pasta.[186] On the other hand, rising gasoline prices will less likely lead people to purchase bicycles.[187] Goods

---

178 *See id.* at 24 ("We may use the model of consumer choice of the previous sections to derive a relationship between the price of a good and the amount of that good in a consumer's optimal bundle.").

179 *See* VARIAN, *supra* note 107, at 35 (providing the axioms of consumer theory).

180 *See* JOHN H. KAGEL ET AL., ECONOMIC CHOICE THEORY: AN EXPERIMENTAL ANALYSIS OF ANIMAL BEHAVIOR 2 (1995) ("[T]he fact that, when put to the test, rats and pigeons conform to elementary principles of economic theory provides rather striking support for the theory and, indirectly, refutes the argument that the theory cannot be extended to nonmarket behavior . . . ."). *See also* LEVITT & DUBNER, *supra* note 96, at 212–13 (describing how one economist taught capuchin monkeys to use money and found that these monkeys also obeyed the most basic law of economics).

181 *See* COOTER & ULEN, *supra* note 25, at 25 (stating that derivation of the demand relationship requires that all other variables be held constant).

182 *See* VARIAN, *supra* note 107, at 95–96 (stating that economists study how demand changes in response to changes in the economic environment).

183 *Cf.* GREENE, *supra* note 153, at 130 (using data from the U.S. gasoline market as an example of instability in a market with changing conditions).

184 *See* VARIAN, *supra* note 107, at 112 ("The idea is that . . . the consumer *substitutes* away from the more expensive good to the less expensive good.").

185 *See* COOTER & ULEN, *supra* note 25, at 25–26 ("Generalizing, the most important determinant of the price elasticity of demand for a good is the availability of substitutes.").

186 *Cf. id.* at 26 ("Substitution is easier for narrowly defined goods and harder for broad categories. If the price of cucumbers goes up, switching to peas or carrots is easy . . . .").

187 *Cf.* JOSEPH E. STIGLITZ, PRINCIPLES OF MICROECONOMICS 79 (2d ed. 1997) (explaining that when gasoline price shocks occur, individuals gradually respond by replacing their vehicles with more fuel efficient alternatives).

like beef have elastic demand, high response to small changes; and goods like gasoline have inelastic demand, low response to large changes.[188]

A more concrete example of abstraction would be a road map. A road map simplifies the relationship between two places.[189] It leaves out mountains, traffic lights, and buildings,[190] but it is useful to a person travelling by automobile because it allows her to see the relative positions in two dimensions as well as the available routes.[191] Depending on the circumstances, however, a road map might not be useful to a person travelling by bicycle, foot, or boat.[192]

A familiar yet intangible example of abstraction would be the scales of justice.[193] The scales of justice weigh all the evidence on both sides of a case.[194] This is an abstraction since we do not literally decide cases by measuring the physical weight of evidence.[195] Obviously oral testimony and eyewitness evidence cannot be weighed. Nevertheless, the model is widely taught and used to describe the relationship between evidence and outcomes.[196]

The point of this discussion about models is to illustrate how the quality of empirical work is limited by the quality of the models. Where we have a strong model we can impose a great deal of structure on our empirical work and obtain fairly precise conclusions.[197] For example, under the model of gravity, Earth's gravitational pull accelerates an ob-

---

188 *See* COOTER & ULEN, *supra* note 25, at 26 (explaining that goods with many possible substitutes will have more elasticity than goods with few substitutes).

189 *See* CASE & FAIR, *supra* note 174, at 10 (stating that simplifying the world as flat on a map is useful).

190 *Id.* at 11.

191 *Id.*

192 *See* Klock, *Wastefulness*, *supra* note 93, at 190 ("A road map will not be very useful to someone traveling on foot in the wilderness or aiming an intercontinental ballistic missile.").

193 *See id.* at 192 (discussing nonexistent things with useful applications in reality, such as Euclidean lines, imaginary numbers, and the scales of justice).

194 *See* Allegra M. McLeod, *Exporting U.S. Criminal Justice*, 29 YALE L. & POL'Y REV. 83, 117 (2010) ("Presented with two opposing sides to a dispute, the judge or jury weighs conflicting evidence to decide which side should prevail").

195 *See* Klock, *Wastefulness*, *supra* note 93, at 192 n.70 ("The phrase 'weight of the evidence' obviously refers to the relative importance assigned to evidence by the arbiter.").

196 *See id.* at 192 ("In fact the scales of justice are a model that attempts to quantify a complex function depending on qualitative arguments.").

197 *See* GREENE, *supra* note 153, at 3 ("With a sufficiently detailed stochastic structure and adequate data, the analysis will become a matter of deducing the properties of a probability distribution."). Professor Greene further writes:

> The process of econometric analysis departs from the specification of a theoretical relationship. We initially proceed on the optimistic assumption that we can obtain precise measurements on all the variables in a correctly specified model. If the ideal conditions are met at every step, the subsequent analysis will probably be routine. Unfortunately, they rarely are.

Id.

ject in a vacuum at sea level at a rate of 9.8 meters per second per second.[198] Where we have a weak model, however, we are unable to impose much structure on our analysis.[199] This does not mean that we should abandon empirical work.[200] We can still learn something from cataloging, measuring, and summarizing empirical data.[201] Our models of independent, cooperative, and vindictive voting are not as strong as the model of gravity, but they can still provide guidance for empirical work.[202]

Empirical work also involves measurement, which itself can involve abstraction.[203] When we attempt to summarize a multi-dimensional concept into a single number, we abstract away from the underlying reality and distort the information by compressing it into a single unit of measurement.[204] The example given by Epstein and King measures George W. Bush as five feet and ten inches tall.[205] If George W. Bush's height is the only measurement, it ignores a great deal of other information about the man.[206] Similarly, observing that Justice O'Connor tended to be conservative ignores more detailed information about the issues for which she was more conservative and the issues for which she was less conservative.[207]

Although my construction of the conservative outcome variable violates one of the rules of good empirical research espoused by Epstein and King, I can defend it. Epstein and King suggest that human judgment should be avoided.[208] I could have reduced some of the statistical noise

---

198 *See* GERALD HOLTON & STEPHEN G. BRUSH, PHYSICS, THE HUMAN ADVENTURE: FROM COPERNICUS TO EINSTEIN AND BEYOND 113 (3rd ed. 2001).

199 *See* GREENE, *supra* note 153, at 4 ("The theory may make only a rough guess as to the correct functional form, if it makes any at all, and we may be forced to choose from an embarrassingly long menu of possibilities.").

200 *See* Epstein & King, *supra* note 2, at 17–18 (concluding that the state of empirical legal scholarship is poor but that it can be improved with more attention to methodological details and rules of inference).

201 *See id.* at 54 ("[U]sing insights from data is a good way to develop theory . . . .").

202 *See* Landes & Posner, *supra* note 46, at 779–80 (suggesting that an informal model can guide empirical analysis).

203 *See* Epstein & King, *supra* note 2, at 81 ("The key is that we abstract the right dimensions for our purposes, and that we measure enough dimensions of each subject to capture all the parts that are essential to our research question.").

204 *See id.* ("[M]easurement allows us to put many apparently disparate events or subjects on the same dimension . . . .").

205 *Id.*

206 *See id.* ("[E]verything about the object of study is lost except the dimension or dimensions being measured.").

207 *See* Landes & Posner, *supra* note 46, at 782 (providing a table of summary statistics showing Justice O'Connor to be conservative overall, but more conservative in civil liberty cases than in economic regulation cases).

208 *See* Epstein & King, *supra* note 2, at 103 ("A study that gives insufficient information about the process by which the data come [sic] to be observed by the investigator cannot be replicated and thus stands in violation of the rule we articulated [earlier].").

in the measure of conservative disposition by eliminating cases for which the classification as conservative was difficult and instead focusing on the easy cases. Indeed I do this in the analytical section by looking at some of the results using the subset of criminal cases for which the definition of conservative and liberal are more straightforward. Nevertheless, making a decision as to which cases are too close to call would still involve human judgment. However, human judgment is less important in this particular study. The reason that Epstein and King suggest that judgment should be avoided is because it makes it impossible for other researchers scrutinizing the analysis to replicate the data.[209] There are two reasons for replicating the analysis. One is to extend the analysis to a different time period. That is irrelevant because this study focuses only on a period of time during which all the Justices on the Court were constant. There can be no extension to other time periods because no other time period had all nine of these Justices on the Court. The other potential reason to replicate the data is to scrutinize or verify the exact same study.[210] That is also irrelevant in this analysis because I will freely give the data in an Excel spreadsheet to any researcher that requests it. Such researchers are then free to use the data as is, or to take issue with my judgments on conservative outcomes in certain cases, and flag and modify them.

IV. EMPIRICAL ANALYSIS

A. *Descriptive Statistics*

We begin the presentation of the empirical results by reporting some overall measures for the Court: the proportion of cases that were criminal, the proportion that were affirmed, the proportion that had a conservative disposition, the proportion that involved unanimity, and the proportion that had 5–4 split decisions. These results are summarized in Table 1 and show the proportions to be 34%, 29%, 59%, 44%, and 21%, respectively. Given the composition of the Court, the figure of 59% conservative dispositions seems very reasonable. The fact that only 29% of the cases are affirmed suggests that this Court was inclined to review

209 *See id.* at 38 ("Good empirical work adheres to the *replication standard*: another researcher should be able to understand, evaluate, build on, and reproduce the research without any additional information from the author.").

210 *See id.* at 42. The authors write:

> [T]he point of the replication standard is to ensure that a published work stands alone so that readers can consume what it has to offer without any necessary connection with, further information from, or beliefs about the status or reputation of the author. The replication standard keeps empirical inquiry above the level of ad hominem attacks on unquestioning acceptance of arguments by authority figures.

Id.

cases they were more likely to reverse or vacate.[211] The proportion of 5–4 decisions is quite large, but it is not even half the proportion of unanimous decisions.[212] This suggests that the Rehnquist Court was not as divided as portrayed by the media and commentators.[213] This is also not surprising because it is known that commentators and reporters tend to focus on controversy.[214] Writing about controversial decisions is more interesting and more likely to result in successful publication than writing about non-controversial decisions.[215] It is useful to know exactly what proportion of cases was unanimous in a stable Supreme Court in order to understand how the selective media publishing biases affect our perceptions of division within the Court.

---

[211] Professors Epstein, Martin, Quinn, and Segal conduct an interesting empirical analysis of individual Justices' votes to affirm and find similar proportions to the aggregate results. *See generally* Lee Epstein et al., *Circuit Effects: How the Norm of Federal Judicial Experience Biases the Supreme Court*, 157 U. PA. L. REV. 833 (2009). These authors argue that

> Under most theories of judging on the Supreme Court, "reversal" is the more plausible forecast. Scholars who study the hierarchy of justice, for example, have noted that the threat of reversal is the only sanction available to Supreme Court Justices against errant circuit courts. Were the Justices to affirm all their decisions, the threat would lose its credibility.

Id. at 871–72.

[212] In a study covering 1937–2004, Professor Landes and Judge Posner found the proportion of decisions decided by a single vote to be 15.2% and the proportion of unanimous decisions to be 30%. Landes and Posner, *supra* note 46, at 790, 800. So the SRCE does have a higher proportion of single vote majority decisions, but also has a substantially larger increase in the proportion of unanimous decisions. It should be noted that Landes and Posner defined unanimity as 9-0 decisions, whereas I defined unanimity to be zero dissents, but since the number of cases with abstaining Justices is small and the number of those which were unanimous is smaller, there will not be much difference attributable to that.

[213] *See generally* Catherine Crier, *Journalism and the Law*, 56 SYRACUSE L. REV. 387 (2006) (describing biases in reporting on the Supreme Court). The attorney-reporter wrote:

> The news media would better serve its readers if journalists acknowledged that the decisions issued by courts at all levels do not necessarily break down along the narrative lines that serve as a template for political stories.
>
> Ironically, even when horse race reporting is somewhat appropriate, members of the media still do the audience a disservice. In their coverage of Congress, state legislatures, and administrative agencies—the very institutions that create the laws and rules at the core of most legal disputes—journalists often fail to explain the real issues.

Id. at 395.

[214] *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1041 ("Newspapers print interesting stories, not dull ones. Editors do not devote scarce space to articles which have uninteresting results.") (footnote omitted).

[215] *See id.* ("Of all the papers written, only the best, most interesting, most provocative, and most surprising will be selected for publication.").

TABLE 1: ATTRIBUTES OF OPINIONS IN THE DATA

Attribute	**Percent**
Criminal Cases	34%
Circuit Court Affirmed	29%
Conservative Disposition	59%
Unanimous Decision	44%
Five-Four Decision	21%
One or More Justices Abstained	4.7%
More than One Justice Abstained	0.1%

Other researchers developing metrics on how liberal or conservative certain Justices are have thrown out the unanimous decisions from their analysis.[216] The argument is that unanimous decisions do not inform us very well about Justices' leanings.[217] Notwithstanding this argument, there is still value in knowing how many decisions are unanimous, and other researchers have published papers that do not provide any information on this proportion.[218]

Table 2 presents statistics for individual Justices. The Justices are listed in order of seniority. This table displays the participation rate (the percentage of the 920 cases the Justice participated in); the batting average (the number of times the Justice voted with the majority or concurred, divided by the number of cases the Justice participated in); a measure of contrariness (the proportion of the Court's decisions with a sole dissenter accounted for by that Justice); and the correlation of each Justice's voting record with the conservative variable. For the construction of the correlation measure, unanimous cases were omitted.

Six of the Justices missed participating in four or fewer of the 920 cases. Justice Kennedy participated in every one,[219] and Justice Ginsburg only missed a single case.[220] Three of the Justices (Rehnquist, O'Connor, and Breyer) missed ten to twelve of the cases. It should be noted though that Justice Rehnquist only missed one case during the first

[216] Andrew D. Martin & Kevin M. Quinn, *Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999*, 10 POL. ANALYSIS 134, 137 n.3 (2002) ("We exclude unanimous cases because they contribute no information to the likelihood. Including unanimous cases also makes it quite difficult to specify reasonable prior distributions for the case parameters . . . .").

[217] *See id.*

[218] *See, e.g.*, Ho & Quinn, *supra* note 156, at 72–76 (using newly collected data on all non-unanimous cases to analyze ideological voting shifts without disclosing any information about the relative frequency of unanimous decisions).

[219] *See infra* Table 2 (showing 100% participation rate).

[220] The case is FEC v. NRA, 513 U.S. 88 (1994).

nine years of this data.[221] His relatively large number of absences during the last two years was most likely a result of his health and treatment.[222]

TABLE 2: METRICS OF JUSTICES

Justice	Participation Rate	Batting Average	Contrarian Percent	Correlation w/ Conservative
Rehnquist	98.70%	.849	3.75%	.554
Stevens	99.57%	.724	58.75%	-.643
O'Connor	98.80%	.891	1.25%	.285
Scalia	99.67%	.792	10.00%	.540
Kennedy	100%	.908	3.75%	.265
Souter	99.67%	.823	5.00%	-.428
Thomas	99.78%	.798	11.25%	.575
Ginsburg	99.89%	.798	3.75%	-.431
Breyer	98.91%	.808	2.50%	-.341

It will not be surprising that the highest batting averages belong to Justices Kennedy and O'Connor, respectively, due to their reputation as swing votes in close decisions.[223] Justice Rehnquist is next, then Justice Souter, followed by Justice Breyer. Justices Ginsburg and Thomas are tied, with Justices Scalia and Stevens having the lowest batting averages. The standout metric is Justice Stevens' contrariness. Justice Stevens accounted for nearly 59% of the Court's cases with a single dissent. The second most contrary Justice is Justice Thomas, who accounted for 11.25% of solo dissents, less than one fifth the amount of Justice Stevens. At the low end, Justice O'Connor had just a single solo dissent[224] accounting for 1.25% of the total of eighty solo dissents in the data.

The metric of conservative values—the correlation of voting with conservative dispositions in non-unanimous decisions—ranks the Justices in order from most conservative to least as: Thomas, Rehnquist, Scalia, O'Connor, Kennedy, Breyer, Souter, Ginsburg, and Stevens. This ordering is nearly the same as that produced by Landes and Posner,

221 The case is Vey v. Clinton, 520 U.S. 937 (1997). This was actually a motion to bring a case against President Clinton without an attorney, and was denied with one Justice dissenting.

222 *Cf. Our Turn: The Case for Limiting Tenure on High Court*, SAN ANTONIO EXPRESS-NEWS, July 31, 2005, at 2H ("Health problems prevented Chief Justice William Rehnquist, now 80 years old with 33 years on the high court, from being present for oral arguments during the recently completed session.").

223 *See* Andrew D. Martin et al., *The Median Justice on the United States Supreme Court*, 83 N.C. L. REV. 1275, 1279 (2005) (identifying Justice O'Connor as a median Justice); Landes & Posner, *supra* note 46, at 802 (identifying Justice Kennedy as a median Justice).

224 The case is *Lebron v. Nat'l Railroad Passenger Corp.*, 513 U.S. 374 (1995).

except that Breyer and Souter are reversed and Ginsburg and Stevens are reversed.[225] The magnitude of the differences in the measurements for Breyer and Souter is trivial, but the magnitude is slightly more noticeable for Ginsburg and Stevens.

Due to the subjectivity involved in classifying case dispositions as conservative, the correlations of the Justices' votes with conservative outcomes is replicated with the subset of criminal cases which are classified by a tighter set of rules. This serves as a check on the robustness of the rank ordering of the conservativeness of the justices. These results are reported in Table 3 which lists the Justices in order from most conservative to least conservative based on the rankings obtained from Table 2. For comparison and reader convenience, Table 3 also provides the Landes-Posner (L-P) metric of conservative voting. It should be noted that L-P measures the proportion of conservative votes in non-unanimous cases that are bounded by zero and one, with a value of 0.5 representing neutrality.[226] The correlations are bounded by negative one and positive one, with a value of zero representing neutrality. Table 3 clearly shows that restricting the analysis to the subset of criminal cases does not change the ranking order of any of the Justices from that found in Table 2. This provides some indication of robustness in the classification of dispositions as conservative.

This is an opportunity to expose another flaw in prior literature on Supreme Court voting. Numerous researchers have calculated metrics for ordering Justices from more conservative to more liberal.[227] For example, Landes and Posner constructed a measure of percentage of votes which were conservative for forty-three Supreme Court Justices' voting records between 1937 and 2006, and then ordered the Justices.[228] There are some obvious qualifications on the rankings that Landes and Posner fairly observe.[229] For example, what it means to be conservative has changed over time, and so it is difficult to compare Justices that served seventy years apart.[230] What is absent from their analysis, however, is any discussion of measurement error and confidence intervals.[231] For example, ranking Justice Thomas as more conservative than Justice Rehnquist because Thomas' score of 0.822 is greater than Rehnquist's

[225] Landes & Posner, *supra* note 46, Table 3 at 782–83.

[226] *See id.*

[227] *See, e.g.*, Jeffrey A. Segal et al., *Ideological Values and the Votes of U.S. Supreme Court Justices Revisited*, 57 J. POL. 812, 815–16 (1995) (providing a table with ideological scores for Supreme Court Justices).

[228] Landes & Posner, *supra* note 46, Table 3 at 782–83.

[229] *See id.* at 781 ("[S]ome of the specific rankings cannot be taken seriously.").

[230] *See id.*

[231] *See id.* at 782–83 (giving point estimates of conservatism without confidence intervals or comparable information). *See also* Segal et al., *supra* note 227, at 816 (giving point estimates of ideological values without confidence intervals or comparable information).

score of 0.815, without providing information or even a discussion about the margin of error, is misleading.[232]

There is a philosophical issue here regarding whether we treat the record of Supreme Court votes as the complete population of votes or as an observed sample (a subset of all cases that could have been voted on).[233] Arguably, the population of Justices' votes consists of all cases that the Court could possibly vote on, and those it did vote on are just a sample drawn from the population of possible cases.[234] In this framework, the metrics calculated by Landes and Posner must be considered *estimates*, rather than parameters.[235] Parameters are descriptive measures of a population, whereas statistics are descriptive measures of a sample that have known and desirable properties (in the case of good statistics) relative to the underlying parameters.[236] Estimates involve estimation error and therefore have margins of error of given size with certain probabilities.[237] To properly infer that Thomas is more conservative than Rehnquist, we need to know that a difference of 0.007 in their conservative scores is *statistically significant*.[238]

Statistical significance is a widely used term that is not well understood in the community of legal scholars.[239] Essentially, for estimates to be statistically significantly different, the discrepancies between them must be large enough to be discernible from what might reasonably oc-

---

[232] Landes & Posner, *supra* note 46, at 782 (ranking Justice Thomas with a score of .822 and Justice Rehnquist with a score of .815).

[233] *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1015 (explaining the difference between a population and a sample).

[234] *See* Thompson & Wachtell, *supra* note 159, at 240–41 (describing the small proportion of cases filed with the Supreme Court that are actually granted a hearing).

[235] *See* FINKELSTEIN & LEVIN, *supra* note 22, at 3 (explaining that attributes of a sample can be useful estimators of population attributes).

[236] *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1016–17 (discussing poor statistics, accurate statistics, and desirable properties of statistics).

[237] *Cf.* FINKELSTEIN & LEVIN, *supra* note 22, at 256 ("The statistician's reason for preferring a random sample . . . is to be able to make probabilistic statements . . . .").

[238] *Cf.* Epstein & King, *supra* note 2, at 98 (complaining that quantitative legal research scholars do not document the procedures they use to obtain their estimates with enough information for the readers to assess the precision of the estimates).

[239] *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1008 ("[C]ommentators and reporters frequently give too much weight to statistics and treat them as actual facts rather than mere estimates which might not be valid or reliable for inferential reasoning."). Finkelstein and Levin make the following complaint:

> Frequently, statistical presentations in litigation are made not by statisticians but by experts from other disciplines, by lawyers who know a little, or by the court itself. This free-wheeling approach distinguishes statistical learning from most other expertise received by the courts and undoubtedly has increased the incidence of models with inappropriate assumptions, or just plain statistical error.

FINKELSTEIN & LEVIN, *supra* note 22, at x. As one example, they give a detailed exposition of numerous flaws in a statistical study embraced by the Supreme Court to conclude that six person juries are as reliable as twelve person juries. *See id.* at 109–10.

cur by chance.[240] How large is large enough depends on three factors: the sample size, the true variation in the population for the underlying variable of interest, and the chosen level of significance.[241] The chosen level of significance refers to the percentage of occurrences considered reasonable to make a Type I mistake—rejecting a correct hypothesis.[242] In order to explain these concepts, I will illustrate with an example.

Suppose that there are two candidates for office, A and B. Let us assume for purposes of simplified calculations that in the population of voters who actually vote on election day, each candidate gets exactly fifty percent of the vote. If in fact we knew this information, there would be no point in conducting a pre-election poll, but we are putting ourselves in the position of Greek gods residing on Mt. Olympus, with a perspective that enables us to see everything.[243] We then ask what inferences could the mere mortals below make when *estimating* the proportion of votes that candidate A will receive from a random sample?

Suppose a pollster randomly selected two people. If each person has a fifty percent chance of supporting candidate A, there are three possible results: both people would support candidate A with a probability of 25%, both people would support candidate B with a probability of 25%, and one would support each 50% of the time. With a sample of two, we would correctly estimate the true proportion of votes 50% of the time. This is not acceptable. Suppose we increase the sample size to four. Now the chances of finding 100% support for either A or B goes from one in two to one in eight. We are much more likely to obtain an estimated value close to 0.5.[244] In fact, the size of our margin of error is inversely proportional to the square root of the sample size.[245] For example, a poll with a sample of ten thousand voters will have a margin of error equal to one-tenth of the margin of error for a poll with a sample of one hundred voters.[246]

Statisticians have a powerful tool called the Central Limit Theorem which tells them that a linear combination of a large number of random

---

240 *See* DAVID R. ANDERSON ET AL., ESSENTIALS OF STATISTICS FOR BUSINESS AND ECONOMICS 226 (abbreviated 4th ed. 2007) (explaining that statistical significance at the á level means that the discrepancy is large enough that if in fact the true difference were zero, a discrepancy of that magnitude would only occur with a probability of á) [hereinafter ANDERSON ET AL., ESSENTIALS].

241 *See id.* at 237 (giving the formula for the size of confidence interval, which just depends on three variables: significance level, variability of the population, and sample size).

242 *See id.* at 225–26 (explaining the meaning of Type I error and significance level).

243 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1024 n.116 (explaining the origin of the term "Olympian knowledge").

244 *See id.* at 1018 ("The probability of getting an estimate of a given size error approaches zero as the sample size increases . . . .").

245 *See* ANDERSON ET AL., ESSENTIALS, *supra* note 240, at 237 (showing that the size of the margin of error is proportional to 1/vn where n is the sample size).

246 $1/10{,}000^{0.5}$=0.1/$100^{0.5}$=0.01.

variables will have a very specific probability distribution known as the normal distribution.[247] This enables the statistician to know exactly what the probabilities are for an estimated proportion from a sample of a given size differing from the true proportion by a given amount.[248] In our example where 50% of voters will vote for candidate A, approximately 95% of all randomly chosen samples of four hundred voters would result in an estimated support level between 45% and 55%.[249] Thus, if we estimated candidate A's support at 48%, we would not consider that significantly different from 50% because a true support level of 50% would generate estimates between 45% and 55% ninety-five percent of the time. If we conducted this poll and got an estimated support level of 44%, however, we know that this large of a deviation from 50% would happen by random chance less than one in twenty times and we might consider that sufficiently small odds to conclude that the support is not at the 50% level.[250]

It is not too surprising that discussion of confidence intervals can be left out of research on voting records since the information required to makes such assessments is normally omitted from polling results.[251] For example, it is common to report that a poll has a margin of error of three percent, or that a poll has a margin of error of four percent.[252] Such reporting is incomplete without also providing the level of significance associated with the margin of error.[253] Any given poll can correctly be said to have any arbitrarily chosen margin of error by varying the level of significance associated with the margin of error.[254]

Hypothesis testing in statistics involves specifying a hypothesis, calculating a test statistic from a random sample of data, and then either

---

247 *See* GREENE, *supra* note 153, at 910 ("[T]he theorem states that sums of random variables, regardless of their form, will tend to be normally distributed. . . . *It requires, essentially, only that the mean be a mixture of many random variables, none of which is large compared with their sum.*").

248 *See* FINKELSTEIN & LEVIN, *supra* note 22, at 113 (explaining how the probabilities of deviant values are calculated using the normal distribution).

249 *See id.* at 171 (giving the formula for the confidence interval as $p \pm 1.96(p(1-p)/n)^{0.5}$ which in the example is $.5 \pm 1.96\times(.5\times.5/400)^{0.5}$.

250 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1019 (describing the process of rejecting a hypothesis with 95% confidence).

251 *See id.* at 1021 n.106 (stating that it is common practice to report poll results without reporting the associated level of confidence).

252 *See id.* ("It is common practice to report that a poll has a margin of error of plus or minus 3% . . . .").

253 *See id.* at 1021 ("[S]tatements about a poll's margin of error without reporting the chosen significance level are uninformative or meaningless.").

254 *See id.* at 1021 n.106 ("Since the margin of error can always be decreased (or increased) by increasing (or decreasing) the significance level, we have no way of knowing how reliable or accurate the poll really is unless the significance level is also disclosed with the margin of error.")

accepting or rejecting the hypothesis.[255] There are four possible outcomes in hypothesis testing: correctly accepting a true null hypothesis,[256] correctly rejecting a false null hypothesis, incorrectly rejecting a true null hypothesis, and incorrectly accepting a false null hypothesis.[257] The first of the two incorrect possibilities is called a Type I error and the second a Type II error.[258] Given that we presume criminal defendants are innocent, we can think of convicting an innocent man as a Type I error and acquitting a guilty man as a Type II error.[259] Note that the probabilities of Type I and Type II errors are not independent of each other.[260] If I always reject the null hypothesis I can never make a Type II error, and if I always accept the null hypothesis I can never make a Type I error.[261] So the smaller I set the probability of a Type I error, the larger will be the probability of committing a Type II error.[262] Statisticians typically set the probability of a Type I error at either 10%, 5%, or 1%.[263] This is what they refer to as the level of significance.[264] At a 5% level of significance, my test statistic will incorrectly reject a true null hypothesis 5% of the time, if the sampling is done correctly.[265]

The relationship between the margin of error and the significance level is such that a larger level of significance (10% being larger than 5%, being larger than 1%), the smaller the margin of error.[266] In other words, if I am comfortable making more Type I errors, I do not need to see as large of a difference between the statistical estimate and the hypothesized value to conclude they are different.[267] To be more concrete,

---

255 *See* DAVID R. ANDERSON ET AL., STATISTICS FOR BUSINESS AND ECONOMICS 314–15 (6th ed. 1996).

256 The null hypothesis is the hypothesis that is being formally tested. We assume it is true and then determine whether the observed data are unlikely to have been generated under that assumption. If so, we reject the null hypothesis; otherwise we accept it. *See* ANDERSON ET AL., ESSENTIALS, *supra* note 240, at 223 (explaining how to construct and test a null hypothesis).

257 *See id.* at 225.

258 *Id.*

259 *See* Ballew v. Georgia, 435 U.S. 223, 234 (1978).

260 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1020 ("[W]e can exert some control over the error rates, though we cannot independently control each error rate.").

261 *See id.* at 1020.

262 *See id.* at 1020–21 ("There is a trade-off in that a lower incidence of one type of error translates into a higher incidence for the other type.").

263 *See* ANDERSON ET AL., ESSENTIALS, *supra* note 240, at 200 (showing the three most commonly used significance levels to be 10%, 5%, and 1%).

264 *Id.*

265 *See id.*

266 *See* PAUL NEWBOLD ET AL., STATISTICS FOR BUSINESS AND ECONOMICS 268–69 (5th ed. 2003) ("[I]f the confidence level (1-α) is decreased, the margin of error will be reduced. For example, a 95% confidence interval will be shorter than a 99% confidence interval based on the same information.").

267 *See id.* at 269 (cautioning that reducing the size of the margin of error increases the probability that the true value lies outside the interval estimate).

a given poll that asserts a margin of error of plus or minus 4% at a significance level of 5% could be correctly reported as having a margin of error of plus or minus 5.2% at a significance level of 1%, and as having a margin of error of plus or minus 3.36% at a significance level of 10%.[268] This is why reporting a margin of error without reporting the significance level is a bad practice—it does not provide all of the information required to ascertain the true accuracy of the poll.[269]

In the context of the current data set, I suggest that the batting average for each Justice can be viewed as an estimate of their true propensity to vote with the majority. The estimate based on the observed cases is noisy because the sample is finite, and some of the differences across Justices should be considered too small to be statistically significant.[270] We would like to have some idea as to how large the differences need to be in order to be considered statistically different. One way to do this is to test the hypothesis of no difference for each Justice; however, there is a statistical problem with that approach. Standard hypothesis testing procedures assume that the observations are independent, meaning, for example, that Justice Scalia's votes do not affect the probabilities of Justice Thomas' votes.[271] So instead of calculating test statistics, I construct confidence intervals. This approach assumes that the cases are randomly selected, but not that the votes across Justices are independent.[272]

Table 3 displays the lower limit and the upper limit of the ninety-five percent confidence intervals for each Justice's batting average. Seeing how wide these intervals are conveys a sense of the precision, or imprecision, of the point estimates for the batting averages. Furthermore, the degree to which the confidence intervals overlap or remain distinctive conveys a sense of how different or similar the Justices' batting averages are. The dissemination of this type of information is extremely valuable in empirical research, and it is frequently missing in the empirical legal scholarship, which often just displays an ordering with

---

268 This follows from the fact that the 10%, 5%, and 1% significance intervals are proportionate to $1.645\sigma/\sqrt{n}$, $1.96\sigma/\sqrt{n}$, $2.576\sigma/\sqrt{n}$, respectively. *See* ANDERSON ET AL., ESSENTIALS, *supra* note 240, at 200 (providing a table that shows the factors of proportionality for each confidence and significance level).

269 *See* Klock, *Finding Random Coincidences*, *supra* note 10, at 1021 n.106.

270 For example, Justice Breyer's estimated propensity to vote with the majority of 0.808 is larger than Justice Ginsburg's estimated propensity of 0.798, but the estimates are not statistically discernible for us to conclude that Justice Ginsburg is truly less likely to vote with the majority than Justice Breyer.

271 *Cf.* NEWBOLD ET AL., *supra* note 266, at 346 (giving the procedure for testing differences between two proportions using independent random samples).

272 *See* FINKELSTEIN & LEVIN, *supra* note 22, at 114 (informing the reader that as long as the observations are independent, the estimated average will have a normal probability distribution).

point estimates without any information as to the accuracy or variability of the point estimates.[273]

TABLE 3: CORRELATION FREQUENCY ANALYSIS OF 5-4 DECISIONS

Justice	**Full Sample**	**Criminal Subsample**	**Landes-Posner Metric**
Thomas	.575	.660	.822
Rehnquist	.554	.653	.815
Scalia	.540	.559	.757
O'Connor	.285	.325	.680
Kennedy	.265	.250	.647
Breyer	-.341	-.351	.372
Souter	-.428	-.459	.374
Ginsburg	-.431	-.467	.312
Stevens	-.643	-.720	.341

In Table 4, we can see that the top four batting averages all overlap to a point. Justice Souter, with the fourth highest estimated batting average, has an upper limit on his confidence interval equal to the lower limit of Justice Kennedy's confidence interval. We also see that the six lowest batting average confidence intervals involve overlap. From this we can conclude that most of the differences across Justices on this measure are not statistically significant. However, a few are. Justice Stevens' average is clearly discernible in statistical terms from Justices Kennedy's, O'Connor's, and Rehnquist's. Justices Scalia, Thomas, and Ginsburg are clearly different from Justices Kennedy and O'Connor. Finally, Justice Breyer's batting average is clearly discernible from Justice Kennedy's. There are thirty-six different pairs of Justices that can be formed from the set of nine, and only ten of the pairs have non-overlapping confidence intervals.[274]

---

[273] *See, e.g.*, Segal et al., *supra* note 227, at 816 (providing a table with ideological scores for Supreme Court Justices).

[274] This is calculated as $9!/(7!\times2!)=36$. *See* FINKELSTEIN & LEVIN, *supra* note 22, at 44 (giving and explaining the formula for counting the number of possible unique subsets of a given size from a group).

TABLE 4: 95% CONFIDENCE INTERVALS FOR BATTING AVERAGES

Justice	**Lower Bound**	**Upper Bound**
Kennedy	87.14%	94.46%
O'Connor	85.13%	93.07%
Rehnquist	80.34%	89.46%
Souter	77.46%	87.14%
Breyer	75.78%	85.82%
Ginsburg	74.71%	84.89%
Thomas	74.71%	84.89%
Scalia	74.05%	84.35%
Stevens	66.83%	78.07%

B. Analysis of 5-4 Decisions

Table 5 provides insight into the cases where the Court is divided 5–4. Elementary counting techniques reveal that there are 126 different possible combinations of five Justices from a group of nine.[275] We have 196 observations on such divisions in the Court. If we randomly assigned cases to each of the 126 different possible combinations, we would expect most cells to have one or two observations. The average value is 196/126, which is 1.5556. Clearly the actual assignments are not random. Based on Table 5, we can see that 74% of these close decisions fell into one of three cells—the five conservatives voting together or one of the two moderate conservatives voting with the four liberals. Table 5 enumerates fourteen specific combinations of the 126 possible combinations. It shows incidences of unusual coalitions, such as each of the three most conservative Justices voting with the four liberals, or each of the liberal Justices voting with the most conservative Justices. Table 5 also reveals, however, that most of the possible coalitions never occurred.

[275] This is calculated as $9!/(5!\times4!)=126$.

TABLE 5: FREQUENCY ANALYSIS OF 5-4 DECISIONS

Coalition	**Frequency** (N=196)
Rehnquist-Scalia-Thomas-O'Connor-Kennedy	89
O'Connor swing vote with 4 liberals	28
Kennedy swing vote with 4 liberals	24
Breyer votes with 3 most conservative and 1 moderate	2
Souter votes with 3 most conservative and 1 moderate	5
Ginsburg votes with 3 most conservative and 1 moderate	4
Stevens votes with 3 most conservative and 1 moderate	2
Rehnquist votes with 4 liberals	2
Thomas votes with 4 liberals	3
Scalia votes with 4 liberals	1
Breyer dissents with 3 most conservative	1
Souter dissents with 3 most conservative	2
Ginsberg dissents with 3 most conservative	2
Stevens dissents with 3 most conservative	1

By adding the frequencies of the fourteen cells in Table 5, we can account for 166 of 196 cases. This leaves only thirty remaining cases, which indicates that there are at most thirty additional combinations or a maximum of forty-five combinations out of the 126 possible. Clearly most of the possible combinations never happened.

Using more detailed information about the cases unaccounted for by Table 5, we can construct a chi-square test statistic to test the hypothesis that the assignments were random.[276] The data for the thirty cases unaccounted for in Table 5 reveals that there are only eighteen additional combinations. Twelve of these combinations are unique, two of them had two occurrences, two of them had three occurrences, and one had five occurrences. Interestingly the combination that occurs five times is the coalition of Scalia, Thomas, Stevens, Souter, and Ginsburg. It is too unwieldy to create a table of all 126 possible combinations, or to list the ninety-four combinations that never occurred, but some examples can be given. The coalition of Thomas, Stevens, Ginsburg, O'Connor, and Kennedy never occurs in the 196 5–4 split decisions. Neither does the coalition of Thomas, Souter, Ginsberg, O'Connor, and Kennedy.

The chi-square test formally tests the null hypothesis that the combinations of voting coalitions in 5–4 cases are randomly dispersed.[277] If

[276] *See id.* at 157–62 (explaining a chi-square test).
[277] *See id.*

the null hypothesis were true, the sum of the squared values of the observed frequency of each cell's coalition minus the cell's expected value under the random assignment hypothesis divided by the cell's expected value would have a chi-square distribution with 125 degrees of freedom.[278] Expressed as an equation: Σ $(f-1.5556)^2/1.5556$ is our test statistic where f equals the observed cell frequency.[279] In this data set, the observed value of the chi-square test statistic is 5,862.26. The critical value for a chi-square with 125 degrees of freedom at 5% significance is 228.58, and at 1% significance is 243.86.[280] The odds of observing the distribution of voting coalitions that we see under the random assignment hypothesis are astronomical.[281] Of course, this is not surprising. We know that certain Justices tend to vote together on controversial issues, but it is still informative to have some measure of the magnitude of the departure from statistical independence in the voting of the Justices.

C. Logit Regressions

The next form of analysis is the most advanced model presented. Logit regression is used to model the votes of one Justice as a function of the other eight Justices.[282] Ordinary regression, also known as least-squares because it minimizes the sum of squared prediction errors, is a technique that is known to most empirical researchers.[283] Ordinary regression is used to estimate relationships when the dependent variable is continuous.[284] In voting models the dependent variable is dichotomous,

---

278 There are 126 cells, but once we know the content of 125 of them we can determine the 126th since the proportions must sum to one. Hence there are only 125 degrees of freedom. *Cf. id.* at 158 (explaining that the sum of n independent squared standard normal random variables has a χ^2 distribution with n degrees of freedom).

279 *See id.* at 157 (stating the formula in words).

280 These critical values are derived from Shazam software. *See generally* WHITE, *supra* note 6, at 317–22 (explaining the calculation of critical values for common distributions).

281 According to Shazam software, the probability of getting a test statistic this large if the null hypothesis is true is about 0.33×10^{-307}. *See id.* (explaining the calculation of probabilities for common distributions).

282 *See generally* FINKELSTEIN & LEVIN, *supra* note 22, at 458–61 (explaining logit regression).

283 *See, e.g.*, Leandra Lederman & Warren B. Hrung, *Do Attorneys Do Their Clients Justice? An Empirical Study of Lawyers' Effects on Tax Court Litigation Outcomes*, 41 WAKE FOREST. L. REV. 1235, 1285 (2006) (using ordinary least squares); *cf.* FINKELSTEIN & LEVIN, *supra* note 22, at 350 ("Multiple regression is a statistical technique for estimating relationships between variables that has . . . invaded the law. . . . It is now so easy to fit models to data by computer that multiple regression and related techniques are likely to become even more widely used . . . .").

284 This follows from the fact that lines are continuous functions, and regression models assume a linear relationship between variables. *See* GREENE, *supra* note 153, at 10 (providing the assumptions of the linear regression model).

each vote is classified as a one or a zero, for or against.[285] Applying ordinary regression in this situation results in serious estimation problems.[286] Logit regression calculates the logarithm of the odds ratio for a positive response.[287] In this model the anti-log of the coefficients on the explanatory variables represents the odds of the dependent Justice voting with the majority, given the explanatory Justice voted with the majority.[288] The anti-log of a negative number is a value less than one, meaning that a negative coefficient implies that the dependent Justice is less likely to vote with the majority if the explanatory Justice voted with the majority.[289] The anti-log of a positive number is, of course, greater than one, which means that a positive coefficient will increase the odds that the dependent Justice votes with the majority if the explanatory Justice did.[290]

Logit regression assumes that the causation runs in one direction.[291] Therefore, it is necessary to assume that the dependent Justice's vote does not affect the voting of the other Justices. I begin the logit analysis with an example using Justice Thomas' voting record as the dependent variable. Justice Thomas is chosen because of the characterization of him as a Scalia clone, and a loyal apprentice to Justice Scalia.[292] Many

---

[285] *See, e.g.*, Christopher B. Colburn & Sylvia C. Hudgins, *The Influence on Congress by the Thrift Industry*, 20 J. BANKING & FIN. 473, 477 (1996) (explaining that a vote for is assigned a value of one and a vote against is assigned a value of zero).

[286] *See* Leandra Lederman, *Which Cases Go to Trial?: An Empirical Study of Predictors of Failure to Settle*, 49 CASE W. RES. L. REV. 315, 348 n.166 (1999) ("Logistic regression was used because where the dependent variable has a dichotomous outcome (here, trial or settlement), ordinary least squares regression can not [sic] be used because the assumption that the errors are homoskedastic is violated.").

[287] *See* FINKELSTEIN & LEVIN, *supra* note 22, at 458.

[288] *See id.* ("The coefficient is therefore referred to as the log odds ratio . . . and its anti-log as the odds ratio or odds multiplier.").

[289] *See id.* at 458–59 ("For example, in a logistic regression involving success or failure on a test, since the anti-log of -0.693 is 0.5, a coefficient of -0.693 for a protected group implies that the odds on passing for a protected group are one-half the odds on passing for a favored group member.").

[290] *See* Lederman, *supra* note 286, at 350 n.176 ("[I]f the log odds for a particular group is 2.30, then cases with that feature are 2.30 times more likely to go to trial and opinion than cases in the reference group (cases without that feature)."). In the present study, the coefficient of Justice Rehnquist on Justice Thomas is 1.6017, the anti-log of which is 4.96. So Justice Thomas is nearly five times more likely to vote with the majority when Justice Rehnquist voted with the majority. When Justice Scalia votes with the majority, Justice Thomas is nearly sixty times more likely to vote with the majority.

[291] *See* WONNACOTT & WONNACOTT, *supra* note 63, at 135–36 (explaining that a change in the independent variable causes a change in the probability of the response variable).

[292] *See, e.g.*, Angela Onwuachi-Willig, *Just Another Brother on the SCT?: What Justice Clarence Thomas Teaches Us About the Influence of Racial Identity*, 90 IOWA L. REV. 931, 933 (2005) ("Justice Thomas has had his independence as a voter on the bench questioned, with the suggestion that he bases his votes on those of a colleague, Justice Antonin Scalia. Indeed, Justice Thomas has been referred to as 'Scalia's puppet,' 'Scalia's clone,' and even 'Scalia's bitch.'") (footnotes omitted).

commentators have dismissed the role of Justice Thomas on the Court, complaining that he simply votes in tandem with other conservative justices on the bench.[293] Other commentators claim that this characterization is unfair;[294] and indeed, the data in Table 2 could support the claim of unfairness as it shows Justices Scalia and Thomas being the solo dissenters more than twice as often as six other Justices. However, my position is neither to support nor attack the characterization of Justice Thomas, but to use the fact that the characterization exists and is widespread as a justification for assuming his voting to be the dependent variable affected by the other Justices for the purpose of demonstrating the use of a logit regression model. Justice Thomas is also famous for not asking his own questions during oral arguments, which could further justify initially modeling his voting as dependent on the other Justices.[295]

Logit regression was only applied to the cases in which all nine Justices participated, there were 877 such cases, in order to avoid observations with missing data.[296] The first column of Table 6 presents the estimated coefficient for each Justice's influence on Justice Thomas and the associated t-statistic. The t-statistic essentially measures whether the estimated coefficient is statistically discernible from zero.[297] A t-statistic of 1.645 is statistically significant at the ten percent level; 1.96 is significant at the five percent level; and 2.576 is significant at the one percent level.[298] The estimated coefficients for Justices Scalia, Rehnquist, and Breyer are highly significant, and the estimated coefficients for the other five Justices are not near statistical significance. The negative coeffi-

[293] *See* Christopher E. Smith, *Clarence Thomas: A Distinctive Justice*, 28 SETON HALL L. REV. 1, 2–3 (1997) ("Thomas has emerged as a distinctive member of the high court. Thomas has . . . articulated themes that distinguish him from all of the other Justices, including the conservative colleagues who share his preferences in determining case outcomes.") (footnotes omitted).

[294] *See* Nancie G. Marzulla, *The Textualism of Clarence Thomas: Anchoring the Supreme Court's Property Rights Jurisprudence to the Constitution*, 10 AM. U. J. GENDER SOC. POL'Y & L. 351, 353 (2002) ("Some commentators have dismissed the role of Justice Thomas on the Court, complaining that he simply votes in tandem with other conservative justices on the bench.").

[295] *See* David A. Karp, *Why Justice Thomas Should Speak at Oral Argument*, 61 FLA. L. REV. 611, 612–13 (2009) (documenting the low quantity of Justice Thomas' comments during oral arguments).

[296] *See* Lederman, *supra* note 286, at 348 ("Multiple regression requires eliminating any case that does not contain information on all of the independent variables used in a particular run.").

[297] *See* Smith v. Xerox Corp., 196 F.3d 358, 366 (2d Cir. 1999) (explaining that a small t-statistic means the difference between two values is too small to eliminate random chance as an explanation).

[298] ANDERSON ET AL., ESSENTIALS, *supra* note 240, at 200.

cient for Justice Breyer indicates that his decisions influence Justice Thomas to vote on the other side.[299]

TABLE 6: LOGIT REGRESSION RESULTS—INFLUENCE OF JUSTICES ON JUSTICE THOMAS

Justice	Coefficient	t-statistic	Coefficient	t-statistic
Scalia	4.0766	13.636	4.0673	14.123
Rehnquist	1.6017	4.0466	1.8623	5.3456
O'Connor	0.31703	0.72909		
Kennedy	0.41199	0.91073		
Stevens	-0.19143	0.42044		
Souter	-0.63615	1.0522		
Breyer	-2.2871	3.3846	-2.4343	3.8292
Ginsbug	-0.68447	1.2073		

The logit regression was repeated without the insignificant Justices. The estimated coefficients for the three remaining Justices and their associated t-statistics are reported in the last two columns of Table 6. The values do not change very much, and this suggests a finding that Justices Scalia, Rehnquist, and Breyer influence Justice Thomas with respect to the eight variable model and the three variable model.

In fairness to Justice Thomas, logit models were estimated with each of the other eight Justices' voting records as the dependent variable. Note that it is not possible for all nine models to be simultaneously correct because each model assumes for the purposes of statistical inference that the causation runs in one direction. Still, the results contain interesting findings. Table 7 provides a summary of these logit regressions and show how each Justices' voting might be significantly positively affected, significantly negatively affected, or unaffected by the other Justices.

---

299 *See* FINKELSTEIN & LEVIN, *supra* note 22, at 458–59 (providing an interpretation for a negative coefficient in a logit regression).

TABLE 7: SUMMARY OF NINE LOGIT REGRESSIONS

Explanatory Justices to Right; Dependent Justice Below	Thomas	Rehnquist	Scalia	O'Connor	Kennedy	Breyer	Souter	Ginsburg	Stevens
Thomas	—	POS	POS	*insig*	*insig*	NEG	*insig*	*insig*	*insig*
Rehnquist	POS	—	POS	POS	POS	*insig*	*insig*	*insig*	NEG
Scalia	POS	POS	—	*insig*	*insig*	*insig*	*insig*	*insig*	NEG
O'Connor	*insig*	POS	*insig*	—	*insig*	POS	*insig*	NEG	NEG
Kennedy	*insig*	POS	*insig*	*insig*	—	*insig*	*insig*	*insig*	POS
Breyer	NEG	*insig*	*insig*	*insig*	*insig*	—	POS	POS	POS
Souter	*insig*	*insig*	*insig*	*insig*	*insig*	POS	—	POS	POS
Ginsburg	*insig*	*insig*	*insig*	NEG	*insig*	POS	POS	—	POS
Stevens	*insig*	NEG	NEG	NEG	POS	POS	POS	POS	—

There are some statistical similarities between the three most conservative Justices (Thomas, Rehnquist, and Scalia). Justice Thomas is positively influenced by Justices Scalia and Rehnquist while negatively influenced by Justice Breyer. Justices Scalia and Rehnquist are each positively influenced by the other two conservative Justices (O'Connor and Kennedy) and negatively influenced by Justice Stevens. Justice Rehnquist is also positively influenced by Justices Kennedy and O'Connor. Justices O'Connor and Kennedy are both positively influenced by Justice Rehnquist, but not by any other conservative Justice. Justice Stevens has a positive influence on Justice Kennedy, but both Justice Stevens and Justice Ginsburg have a negative influence on Justice O'Connor.

There is a conflict between the two women of the SRCE that is statistically discernible. Both women negatively influence each other and Justice Stevens has an opposing influence on the two women. Justice Ginsburg is positively influenced by Justices Souter and Breyer. Justice Stevens is positively influenced by Justice Kennedy and the three more liberal Justices (Ginsburg, Souter, and Breyer). Justice Stevens is negatively influenced by Justices O'Connor, Rehnquist, and Scalia and unaffected by Justice Thomas. Justice Breyer is negatively influenced by Justice Thomas. Justice Breyer is also positively influenced by Justice O'Connor and each of the three more liberal Justices.

Justice Souter's statistics are interesting. Justice Souter is positively influenced by the three more liberal members of the court (Stevens, Ginsburg, and Breyer), but he is not negatively influenced by anyone. Justice Souter and Justice Kennedy are the only members of the court who are not negatively influenced by another Justice, although Justice

Souter has more positive role models. Justices Souter and Kennedy are also the only members of the court who do not show a negative influence on any other Justice; however, Justice Souter also has a positive effect on more of the other Justices. Justice O'Connor has a negative influence on two Justices and Justice Stevens has a negative influence on three Justices. The other five Justices exert negative influence on one other member of the Court (Thomas, Rehnquist, Scalia, Breyer, and Ginsburg). Justices Rehnquist, Stevens, and Breyer each affect four other Justices positively. Justices Ginsburg and Souter each affect three positively; Justices Kennedy, Thomas and Scalia each affect two positively; and Justice O'Connor only has a positive influence on one (Rehnquist). If net influence is defined to be the number of Justices a Justice influences positively minus the number influenced negatively, Justice O'Connor actually has a negative net influence. The Justices with the largest net influence metric are Rehnquist, Breyer, and Souter, all tied at a net measure of three.

The logit regressions also suggest that Justice Kennedy is, in one sense, the most independent Justice. His voting record is only statistically significantly affected by two other Justices, Rehnquist and Stevens. Both Justices have a positive influence on Justice Kennedy, even though Justice Rehnquist has a negative influence on Justice Stevens and Justice Stevens has no statistically significant effect on Justice Rehnquist.

Table 8 provides the exact values of the statistically significant t-statistics for the logit regressions. All of the information in Table 7 is contained in Table 8, but Table 7 makes visualization of the positive and negative influences easier. Table 8 provides the numerical values to allow readers to get a feel for the magnitudes of the significance levels. Each logit regression was then replicated by removing the statistically insignificant Justices from the regression. In the interest of conserving space, these results are not reported in Tables, but the statistical significance of all the remaining Justices was preserved so the findings are robust with respect to this choice of model specification.

TABLE 8: SIGNIFICANT T-STATISTICS FROM NINE LOGIT REGRESSIONS

Explanatory Justices to Right; Dependent Justice Below	Thomas	Rehnquist	Scalia	O'Connor	Kennedy	Breyer	Souter	Ginsburg	Stevens
Thomas	—	4.05	13.64	*insig*	*insig*	-3.38	*insig*	*insig*	*insig*
Rehnquist	4.14	—	3.54	5.14	6.45	*insig*	*insig*	*insig*	-4.05
Scalia	13.76	3.18	—	*insig*	*insig*	*insig*	*insig*	*insig*	-2.21
O'Connor	*Insig*	5.17	*insig*	—	*insig*	6.76	*insig*	-2.62	-2.45
Kennedy	*Insig*	6.52	*insig*	*insig*	—	*insig*	*insig*	*insig*	3.36
Breyer	-3.73	*insig*	*insig*	*insig*	*insig*	—	2.85	6.80	4.89
Souter	*Insig*	*insig*	*insig*	*insig*	*insig*	3.57	—	8.47	5.30
Ginsburg	*Insig*	*insig*	*insig*	-2.19	*insig*	6.86	8.42	—	6.63
Stevens	*insig*	-4.33	-2.51	-2.58	3.73	5.26	4.90	6.47	—

It may be true that Justice Thomas is the least independent thinker. By all measures of fit, Justice Thomas' voting was the most predictable based on the other Justices' votes. The logit regression correctly predicted his vote 93% of the time. The Cragg-Uhler R-squared measure, found in Table 9, contains a measure of goodness of fit for the logit regressions for the nine Justices.[300] The Justices have been listed in the order of best fit, and the results demonstrate that the model fits Justice Thomas better than any other Justice.[301]

TABLE 9: CRAGG-UHLER MEASURE OF FIT FOR LOGIT REGRESSION MODELS

Dependent Justice	**R-Squared**
Thomas	.7123
Scalia	.6790
Ginsburg	.6568
Souter	.5946
Breyer	.5942
Stevens	.5596
Rehnquist	.5590
Kennedy	.2963
O'Connor	.2923

300 *See* John G. Cragg & Russell S. Uhler, *The Demand for Automobiles*, 3 CANADIAN J. ECON. 386, 400 n.20 (1970) (defining an R-squared measure for logit regressions).

301 The larger the R-square, the better the fit. An R-square of 1.0 represents a perfect fit. FINKELSTEIN & LEVIN, *supra* note 22, at 369.

D. *Tests for Structural Change*

It is reasonable to suspect that voting behavior changes over time.[302] The Chow test is a statistical test for a change in regime that can be used to test the null hypothesis that voting behavior is constant over time against the alternative hypothesis that a structural change has occurred.[303] The procedure involves splitting the sample into two time periods and estimating three regressions—one regression for each subperiod and a third regression which combines both periods.[304] If the sum of squared residuals does not change very much when the two periods are estimated individually, then the null hypothesis of no structural change will be accepted.[305] If the sum of squared residuals does change by a large amount that is statistically discernible, then there is evidence of a structural change.[306]

There is some evidence that the power (the probability of not making a Type II error) of the test procedure can be increased by omitting some of the observations in the middle.[307] For these tests cases from calendar year 1999 were excluded and the subsamples of decisions prior to 1999 and decisions subsequent to 1999 were used. I then calculated a test statistic for the test of structural change for each of the nine logit models. These statistics have an F distribution with 9 degrees of freedom in the numerator (the number of regressors—each Justice plus an intercept) and 788 degrees of freedom in the denominator (806 observations excluding 1999 and decisions with missing Justice votes less the 18 estimated coefficients—one for each Justice plus an intercept in two separate subsamples).[308] Therefore, the critical values of the test statistic at significance levels of ten percent, five percent, and one percent are 1.648, 1.900, and 2.432 respectively.[309] The results of the tests are reported in Table 10.

---

[302] *See, e.g.*, Landes & Posner, *supra* note 46, at 789 ("It has been suggested that a Justice's judicial ideology might vary over his tenure . . . .").

[303] *See* GREENE, *supra* note 153, at 130 (explaining the Chow test for structural change).

[304] *See id.* at 130–31 (describing the test procedure).

[305] *See id.* at 131 (explaining that the test statistic is derived from the change in the sum of squared residuals between the restricted and unrestricted regressions).

[306] *See* GUJARATI, *supra* note 56, at 264 (stating that the hypothesis of no structural change should be rejected if the test statistic is sufficiently large).

[307] *Cf.* WHITE, *supra* note 6, at 173 (giving the statistician the option to exclude observations from the middle of the sample in conducting a Chow test).

[308] *See* GUJARATI, *supra* note 56, at 263–64 (explaining how to calculate degrees of freedom for a Chow test). *See generally* NEWBOLD ET AL., *supra* note 266, at 242–43 (explaining the degrees of freedom concept).

[309] These critical values are derived from Shazam software. *See generally* WHITE, *supra* note 6, at 317–22 (explaining the calculation of critical values for common distributions).

TABLE 10: CHOW TESTS FOR STRUCTURAL CHANGE

Dependent Variable	Test Statistic	p-Value
Thomas	2.65	0.0049827
Rehnquist	6.07	0.00000003
Scalia	0.78	0.63505
O'Connor	5.51	0.0000002
Kennedy	2.63	0.0053124
Breyer	2.03	0.033574
Souter	5.16	0.0000008
Ginsburg	6.44	0.000000007
Stevens	3.47	0.0003255

The p-value demonstrates what the significance level would have to be to make the test statistic borderline significant.[310] The lower the p-value the more we are able to reject the null hypothesis of no change with a very low probability of a Type I error.[311] Alternatively, the lower the p-value, the more statistically significant the test statistic is.[312] These tests indicate that only Justice Scalia has no statistically significant change in voting relationships during the SRCE at any conventional level of significance. Of the other eight Justices, we can reject the null hypothesis of no change at a ten percent level of significance (error rate) for all; at a five percent level of significance for all but Justice Breyer; and at a one percent level of significance for Justices Rehnquist, O'Connor, Souter, Stevens, and Ginsberg. The last finding is comparable to the findings of Professor Landes and Judge Posner who reported statistically significant shifts in ideology for Justices Rehnquist, O'Connor, Souter, Stevens, and Ginsberg.[313]

CONCLUSION

Many of the empirical findings are already known and not surprising—Justices Kennedy and O'Connor were frequently the swing voters during the SRCE, and most of the 5–4 decisions split along traditional

---

310 *See* ANDERSON ET AL., ESSENTIALS, *supra* note 240, at 231 ("[T]he p-value is also called the *observed level of significance*.").

311 *See id.* at 229 ("[A] small p-value indicates a sample test statistic that is unusual given the assumption the H_0 is true.").

312 *See* Gross & Syverud, *supra* note 132, at 334 n.48 ("Note that the *smaller* the p-value the *greater* the confidence that the results do not reflect mere chance fluctuations.").

313 Landes & Posner, *supra* note 46, at 790 n.16.

conservative and liberal lines.[314] However, there is incremental value in having measures that reveal the precise degree to which these generalizations are true.[315] Additionally, some of the conventional wisdom that the swing voting Justices are the most important might not be true.[316] The logit models of voting suggest that the Justices with the greatest statistically significant net influence on other Justices were Rehnquist, Souter, and Breyer.

I have endeavored to be somewhat more careful and more thorough in my empirical analysis than I believe other scholars have been. First, this analysis is limited to a stable court so that any Justice's voting record is confined to a period when all other Justices were constant. The Chow tests reveal that even with the composition of the Court held constant, there might be structural changes over time that makes modeling the votes of the Court problematic. Additionally I provide data on the frequency of both unanimous cases and non-unanimous cases. I report confidence intervals that reveal the precision (or imprecision) of some of the metrics. I therefore can report that the batting averages across twenty-six of thirty-six different possible pairs are not statistically discernible.

The chi-square statistic for the null hypothesis that the 5–4 decisions were random reveals that the perceived division in the Stable Rehnquist Court was real and extremely large in statistical terms. The logit regressions also confirm that Justice Thomas is the most consistently predictable member of the Court based on the votes of the others.

The empirical analysis provides many insights about the Stable Rehnquist Court, but it is not capable of determining which model of voting best describes the Court: independent, cooperative, or vindictive. The votes are not independent, but it is unknown whether the cause of the statistical dependence is some exogenous variable outside the model, or whether instead the Justices work cooperatively or vindictively.[317] There is more positive influence between pairs of Justices than negative influence, which can be interpreted as meaning that there is more cooper-

---

[314] *See, e.g.*, Peter B. Rutledge, *Looking Ahead: October Term 2006*, 2005–2006 CATO SUP. CT. REV. 361, 367 (2005–06) ("Since he joined the Court in 1988, Justice Kennedy has shared with Justice O'Connor the power of serving as 'swing justice' on most issues.").

[315] *Cf.* Christopher P. Guzelian et al., *A Quantitative Methodology for Determining the Need for Exposure-Prompted Medical Monitoring*, 79 IND. L.J. 57, 96–97 (2004) (expressing a preference for quantitative measures over qualitative expressions).

[316] *See* Epstein & Jacobi, *supra* note 17, at 40 (2008) ("[I]n theory the median Justice should be quite powerful . . . .").

[317] *See* Ruger et al., *supra* note 8, at 1190–91 (2004) (explaining that observing ideology related correlations in decision making does not mean that other factors are not the cause of the relationships).

ation than retaliation.[318] However, the analysis can only examine relationships between individuals, not between blocks. The methodology employed cannot investigate whether blocks of Justices retaliate against others. Nevertheless, although the data cannot both determine the true model and validate it, it does provide many informative metrics.[319]

Many individuals can look at the same data and come to opposing conclusions.[320] Data that student test scores improved from one year to the next could be used to argue that teaching performance improved.[321] Alternatively, the data could be used to argue that the teachers cheated and gave students the answers to the test questions.[322] There are enough non-traditional coalitions in the data to dispel both notions of vindictive and cooperative voting. Although the voting patterns are clearly not randomly dispersed, one should not expect them to be random because some Justices can clearly be labeled as more or less conservative, and the votes are expected to be correlated as a consequence of the fact that many cases split along conservative and liberal ideologies. Perhaps the empirical fact that should be emphasized is the simplest one—forty-four percent of the decisions during the SRCE were unanimous—a high percentage by twentieth century standards.[323]

---

[318] *See e.g.*, Justice Clarence Thomas, *Remarks from the 100th Arkansas Bar Association Convention*, 51 ARK. L. REV. 651, 653 (1998) (describing cordial daily lunches amongst the Justices and how well they all like each other).

[319] *See* Leamer, *supra* note 62, at 36 (stating that data can reveal some information, but data alone cannot reveal the full relationship between variables).

[320] *See* STEVEN D. LEVITT & STEPHEN J. DUBNER, FREAKONOMICS: A ROGUE ECONOMIST EXPLORES THE HIDDEN SIDE OF EVERYTHING 12 (2005) ("*The conventional wisdom is often wrong*.").

[321] *See id.* at 29 ("A dramatic one-year spike in test scores might initially be attributed to a *good* teacher . . . .").

[322] *See id.* at 27–28 ("But if a teacher *really* wanted to cheat—and make it worth her while—she might collect her students' answer sheets and, in the hour or so before turning them in to be read by an electronic scanner, erase the wrong answers and fill in the correct ones.").

[323] *See* Landes & Posner, *supra* note 46, at 790–91 ("[A]bout 30 percent of the Supreme Court decisions in the 1937-2004 period were decided unanimously . . . . The fraction of unanimous decisions has been trending upward from around 30 percent in the 1960s, and is now in the 40 percent range . . . .").