University of Michigan Journal of Law Reform

Volume 21

1988

The Effectiveness of Measures to Increase Appellate Court Efficiency and Decision Output

Thomas B. Marvell
Court Studies, Inc.

Carlisle E. Moody
College of William and Mary

THE EFFECTIVENESS OF MEASURES TO INCREASE APPELLATE COURT EFFICIENCY AND DECISION OUTPUT†

Thomas B. Marvell* and Carlisle E. Moody**

Appellate court caseloads have increased greatly since World War II, doubling approximately every ten years.[1] Because states created appellate judgeships only one-third as fast as caseloads grew,[2] the judges have faced heavy demands to increase productivity. Most state appellate courts successfully expanded decision output to match the appeals "explosion"; decision growth averaged 116% over the ten-year period from 1974 to 1984, only slightly less than the 123% growth for filings over the corresponding period.[3] Output per judge increased, on average, from fifty-three to eighty-eight decisions in the 1974-1984 period for the forty-five states with available information.[4]

The plight of the appellate courts has resulted in an enormous body of literature recommending procedures for increasing judicial resources and judicial productivity.[5] The literature, however,

† This research was supported by Grant 83-IJ-CX-4046 given by the National Institute of Justice to Court Studies, Inc. The contents do not necessarily reflect the views of the Institute.

* Court Studies, Inc., Williamsburg, Virginia. B.A., Harvard, 1961; J.D., University of Michigan, 1964; Ph.D. (sociology), University of Michigan, 1976.

** Department of Economics, College of William and Mary, Williamsburg, Virginia. B.A., Colby College, 1965; Ph.D. (economics), University of Connecticut, 1970.

1. *See, e.g.*, BUREAU OF JUST. STATISTICS, U.S. DEP'T OF JUST., BULLETIN: 1973-83 TRENDS, THE GROWTH OF APPEALS 2 (1985) (discussing increase in appellate caseload over recent 10-year period); Marvell, *Appellate Court Caseloads: Historical Trends*, 4 APPELLATE CT. ADMIN. REV. 3 (1983). Research in progress by the authors, however, suggests that appellate filing growth slowed greatly during the mid-1980's, especially in civil cases. This trend was forecast, *id.* at 7, because civil caseload trends tend to follow economic trends, and the nation experienced recessions in 1980 and 1982.

2. BUREAU OF JUST. STATISTICS, *supra* note 1, at 3.

3. Marvell, *State Appellate Court Responses to Caseload Growth*, 72 JUDICATURE (1988) (forthcoming) [hereinafter *Responses to Caseload Growth*]. This is based on 39 states' filings for the 10 years from 1973 to 1983 and decisions for the 10 years from 1974 to 1984. Decisions lag a year because courts take about a year to decide appeals. "Decisions" and "filings" are defined *infra* note 42 and text accompanying 56.

4. *Responses to Caseload Growth*, *supra* note 3.

5. *See generally* AMERICAN BAR ASS'N TASK FORCE ON APPELLATE PROCEDURE, EFFICIENCY AND JUSTICE IN APPEALS: METHODS AND SELECTED MATERIALS (1977); P. CARRINGTON, D. MEADOR & M. ROSENBERG, JUSTICE ON APPEAL (1976); COMMISSION ON STAN-

has not provided more than educated guesses about which procedures work best. This Article attempts to fill this void by evaluating the changes that state appellate courts have made to increase decision output. The changes fall into seven categories: (1) adding judges or temporarily assigning trial and retired judges; (2) hiring law clerks and staff attorneys; (3) curtailing opinion practices by deciding cases with unpublished opinions, with unsigned opinions, or without opinions; (4) creating or expanding intermediate courts; (5) reducing the number of judges who participate in each decision; (6) curtailing oral arguments; and (7) using summary judgment procedures. These categories encompass virtually all of the major changes that appellate courts have made to increase productivity in the past two decades.[6] Although questions of productivity and efficiency are never the sole criteria for determining how appellate courts should adapt to rising caseloads, knowledge about which changes actually do increase productivity and efficiency, and by how much, should be valuable information when making such decisions.

This Article will examine the effectiveness of measures commonly employed to increase appellate court productivity. Part I of the Article sets forth some common design problems and explains how the research technique employed in the present study avoids these problems by using a multiple time-series research design. Part II applies this design to state court data. Part II also describes the dependent variable, the number of appeals decided per judge, used in the regression analysis. Part III discusses the results of that analysis—the impact of each change listed above on judicial productivity. The Article, although not advocating the adoption of the discussed efficiency measures, concludes that the failure to enact any type of efficiency measure will cause appellate courts to fall behind in the handling of their caseloads.

DARDS OF JUDICIAL ADMIN., AM. BAR ASS'N, STANDARDS RELATING TO APPELLATE COURTS (1977); R. LEFLAR, INTERNAL OPERATING PROCEDURES OF APPELLATE COURTS (1976); T. MARVELL, BIBLIOGRAPHY: STATE APPELLATE COURT ADAPTATION TO CASELOAD AND DELAY PROBLEMS (1985); T. MARVELL, BIBLIOGRAPHY: STATE APPELLATE COURT WORKLOAD AND DELAY (1979); D. MEADOR, APPELLATE COURTS: STAFF AND PROCESS IN THE CRISIS OF VOLUME (1974); Marvell, *Appellate Capacity and Caseload Growth*, 16 AKRON L. REV. 43 (1982).

6. *See infra* note 45.

I. RESEARCH DESIGN

Empirical research requires good research design. In the same manner that a building cannot be trusted if its structural design is substandard, research findings lack credibility if the research design is faulty. The empirical study of courts, however, differs from building construction in that researchers frequently ignore the generally accepted design standards. As a result, judges and others should not rely on the conclusions of most court research.

The major difficulty in empirical research is differentiating the impact of numerous factors that affect the studied variables. The present research, evaluating changes made in appellate courts, provides a good illustration. One cannot easily distinguish the impact of any one particular change, such as adding staff attorneys, from the impact of other innovations designed to improve efficiency and from the impact of numerous other developments in the courts. Most appellate courts have taken multiple procedural and structural steps to increase productivity. Further, these steps are not one-shot, discrete events; rather, the courts regularly expand the use of particular innovations. For example, courts typically increase attorney staffs incrementally, and they usually phase in new efficiency programs—such as reducing opinion publication—over several years. Courts, moreover, sometimes reverse or eliminate such steps, especially when new judgeships are created.

Productivity is also affected by innumerable events other than the changes designed to improve operations. The most important, as will be explained later, is the volume of filings. Judge turnover and changes in composition of caseloads may also be important. Furthermore, caseload volume can indirectly cause courts to adopt procedures designed to stimulate output—reciprocal causation. In sum, there is a tangled mass of forces operating in each court, and an acceptable research design must present a credible mechanism for distinguishing the impact of particular forces.

A large volume of literature exists that specifies which research designs are acceptable for distinguishing between causal factors.[7] Still, commentators criticize research on the courts and

---

7. *See generally* D. CAMPBELL & J. STANLEY, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH (1963); T. COOK & D. CAMPBELL, QUASI-EXPERIMENTATION: DESIGN & ANALYSIS FOR FIELD SETTINGS (1979). For law and society research, see J. MONAHAN & L. WALKER, SOCIAL SCIENCE IN LAW (1985); Lempert, *Strategies of Research Design in the Legal Impact Study: The Control of Plausible Rival Hypotheses*, 1 LAW & SOC'Y

other aspects of the justice system for using poor research designs.[8] Researchers in this area, in fact, generally ignore the accumulated social science knowledge about how to conduct research: They seldom refer to, or otherwise base their methods on, methodological texts or other authorities, and they infrequently warn readers that research design problems render the results uncertain. From the social science perspective, these omissions are tantamount to professional malpractice. The result is the application of Gresham's law: "Bad research so turns off decision makers that good research is largely ignored."[9]

A. Unacceptable Research Designs

Two research designs that social science methodological writings single out as inadequate are the before-and-after design and the cross-section design. Nevertheless, court researchers frequently use these designs.

1. The before-and-after design— Perhaps the most common, and worst, design for evaluating changes in courts simply compares the situation in the years before the change to the situation in the years after. To study the impact of adding staff attorneys, for example, the before-and-after design compares the number of cases decided per judge in one or a few years before hiring the staff to the number of cases decided per judge during the following few years. Such studies are virtually useless for several reasons.

The first and most obvious problem is that other factors may have caused any increase or decrease in decision output measured. For instance, one cannot determine whether the staff attorneys had an effect unless one concludes that all other likely factors are known, and that they did not influence the results. Such conclusions are rarely justified.

REV. 111 (1966); Lind, Shapard & Cecil, *Methods for Empirical Evaluation of Innovations in the Justice System*, in FEDERAL JUDICIAL CENTER ADVISORY COMM. ON EXPERIMENTATION IN THE LAW, EXPERIMENTATION IN THE LAW 85-121 (1981); Luskin, *Building a Theory of Case-Processing Time*, 62 JUDICATURE 114 (1978).

8. A major review of federally funded criminal justice research, including court research, concluded that the quality of such research was predominantly mediocre, mainly because of poor research designs. COMMITTEE ON RESEARCH ON LAW ENFORCEMENT AND CRIMINAL JUST., NAT'L RESEARCH COUNCIL, UNDERSTANDING CRIME: AN EVALUATION OF THE NATIONAL INSTITUTE OF LAW ENFORCEMENT AND CRIMINAL JUSTICE 65-71, 92-96 (1977). *See also* Luskin, *supra* note 7.

9. Johnson, *Misuses of Applied Social Research*, in THE USE/NONUSE/MISUSE OF APPLIED SOCIAL RESEARCH IN COURTS 39, 41 (M. Saks & C. Baron eds. 1980).

A good example of this type of problem appeared in an evaluation of an Arizona program designed to reduce delays by using volunteer lawyers as temporary judges. Division One of the Arizona Court of Appeals used a large number of lawyer-judges from late 1984 until mid-1986, such that the court was able to increase the effective number of judges by nearly twenty percent.[10] A study of the program[11] found that the time to decision declined substantially in 1985 and 1986, and it attributed this change to the lawyer-judges, as well as to increased efforts by the regular judges.[12] The delay reduction, however, may have had other causes. After several years of sharp increases, civil appeals coincidentally declined when the program began,[13] and another division of the court absorbed much of Division One's caseload, further reducing caseload pressure.[14] Such major reductions in caseloads and backlogs provide an explanation for the delay reduction that is at least as likely as the use of lawyer-judges. Other changes that may have contributed to the delay reduction include the declining use of full opinions in favor of shorter unpublished "memorandum" opinions[15] and the addition of a second law clerk for most judges in 1985.[16]

---

10. A. AIKMAN, M. ELSNER & F. MILLER, FRIENDS OF THE COURT: LAWYERS AS SUPPLEMENTAL JUDICIAL RESOURCES 84-111 (1987).

11. *Id.*

12. *Id.* at 98, 106-07.

13. Attorneys filed 537 civil appeals in 1980, ADMINISTRATIVE OFFICE OF THE COURTS, THE ARIZONA COURTS: 1982 ANNUAL JUDICIAL REPORT 17 (1983) [hereinafter 1982 JUDICIAL REPORT], 605 in 1984, ADMINISTRATIVE OFFICE OF THE COURTS, THE ARIZONA COURTS 1984 JUDICIAL REPORT 7 (1985) [hereinafter 1984 JUDICIAL REPORT], 598 in 1985, ADMINISTRATIVE OFFICE OF THE COURTS, THE ARIZONA COURTS: 1985 JUDICIAL REPORT 12 (1986) [hereinafter 1985 JUDICIAL REPORT], and 563 in 1986, ADMINISTRATIVE OFFICE OF THE COURTS, THE ARIZONA COURTS: 1986 JUDICIAL REPORT 15 (1987) [hereinafter 1986 JUDICIAL REPORT]. A. AIKMAN, M. ELSNER & F. MILLER, *supra* note 10, at 85, erroneously show a five-percent increase in 1985.

14. *See, e.g.*, 1985 JUDICIAL REPORT, *supra* note 13, at 12-13. Arizona's second intermediate court, Division Two, grew from three to six judges in early 1985. *Id.* at 13. Division One's backlog of civil appeals was significantly reduced by transferring a large number of cases from Division One to Division Two. In 1985, there was a net transfer, out of Division One, of 52 cases. Fifty of these cases were transferred to Division Two. *Id.* at 12-13. In 1986, a net 137 cases were transferred out of Division One, and 134 net cases were transferred into Division Two. 1986 JUDICIAL REPORT, *supra* note 13, at 15-16.

15. The court reports show that full opinions declined from 30% of civil decisions in 1984, 1984 JUDICIAL REPORT, *supra* note 13, at 7, to 19% in 1985, 1985 JUDICIAL REPORT, *supra* note 13, at 12, and 25% in 1986, 1986 JUDICIAL REPORT, *supra* note 13, at 15.

16. In 1983, there were 16 law clerks for the 15 judges in both divisions. ADMINISTRATIVE OFFICE OF THE COURTS, THE ARIZONA COURTS: 1983 CASELOAD, FINANCIAL, AND PERSONNEL REPORT (1984). In 1984, there were 21 clerks for 15 judges. ADMINISTRATIVE OFFICE OF THE COURTS, THE ARIZONA COURTS: 1984 CASELOAD, FINANCIAL, AND PERSONNEL REPORT (1985). In 1985, there were 31 clerks for 18 judges. ADMINISTRATIVE OFFICE OF

The before-and-after design also cannot distinguish the effects of innovations from "secular drift," a general long-term trend. Appellate caseloads increased regularly for several decades until the mid-1980's, prompting courts to dispose of more appeals irrespective of any procedural or personnel changes. Thus, a before-and-after study usually shows that an innovation works.[17]

Researchers often compare one measure taken before the innovation to a different measure taken afterwards. For example, a study of summary procedures in the Rhode Island Supreme Court concluded that new summary procedures are efficient largely because the average delay in *cases subjected to the procedures* was less than the delay for *all criminal cases* in the prior year.[18] A frequent mistake is to measure delay in cases filed before and after an innovation, but to exclude cases still pending at the time researchers wish to conclude the research. The pending cases are predominately those filed after the innovations and those with the longest delay, so the sample of cases for the "after" period is skewed because it includes only the cases with less delay.[19]

A final problem is that the innovation may not have an immediate impact because it may take some time for judges and staff to overcome the inertia of the old standard operating procedures and to iron out the bugs in the new procedures.[20] Therefore, the usual procedure of only measuring output soon after the experiment is initiated is not sufficient.

2. *Cross-section studies*— A second unacceptable design is the cross-section survey. For instance, a researcher might study

---

THE COURTS, THE ARIZONA COURTS: 1985 CASELOAD, FINANCIAL, AND PERSONNEL REPORT (1985).

17. *See, e.g.*, J. LUCAS, THE APPELLATE PROCESS AND STAFF RESEARCH ATTORNEYS IN THE ILLINOIS APPELLATE COURT 73-84 (1974). In that study, the researcher concluded that staff attorneys helped courts increase the number of cases decided, because the number went up in the year staff attorneys were introduced. *Id.*

18. *See* Olson & Chapper, *Screening and Tracking Criminal Appeals: The Rhode Island Experience*, 8 JUST. SYS. J. 20 (1983). The cases subjected to summary procedures were the more routine cases, and they were given abbreviated briefing schedules. The study did not give the time to disposition for all criminal appeals in the "after" period. A similar problem was encountered in a study conducted by Chapper and Hanson. *See* Chapper & Hanson, *Expedited Procedures for Appellate Courts: Evidence from California's Third District Court of Appeal*, 42 MD. L. REV. 696 (1983). The study found that delay in cases subjected to a procedure that gave them priority in scheduling was less than a similar group of cases decided the year before but not given scheduling priority. *Id.* at 720. Again, this delay reduction may well have been at the expense of other appeals, but no evidence on that point was presented.

19. *E.g.*, A. AIKMAN, M. ELSNER & F. MILLER, *supra* note 10, at 92, 95-96.

20. Casper & Brereton, *Evaluating Criminal Justice Reforms*, 18 LAW & SOC'Y REV. 121, 132, 138-42 (1984).

several courts, comparing output volume with characteristics of the courts. This type of research is unacceptable for several reasons. First, the sample size—the number of units studied—is often too small to justify conclusions.[21] A fifty-state sample would not be sufficient for a study of appellate courts, especially because a researcher should include a large number of variables to control for other differences between the states.[22] Second, the available variables may not capture some major differences between courts, causing misleading results.[23] Third, causal interpretation is often impossible. Because the data are taken at one point in time, correlations can result because the dependent variable (e.g., decisions per judge) affects the independent variables (factors that are thought to affect decisions per judge). A finding that courts with more staff attorneys are more productive, for example, cannot justify a conclusion that the larger staff enhances productivity; a likely alternative interpretation is that courts with high caseloads (and thus more decisions) per judge try to relieve workload pressures by hiring more staff.[24]

Numerous court studies use the case as the unit of analysis.[25] This strategy overcomes the problem of small sample size, but the research usually fails because almost all relationships between the dependent variable (e.g., decision output or delay), and procedural features of the cases are causally ambiguous. A finding that cases decided without published opinion are decided more quickly than other cases, for example, does not mean that the procedure speeds opinions; more likely, cases selected

---

21. Court research provides numerous examples of cross-section studies that cannot provide meaningful information because they use ludicrously small samples, often outnumbered by the number of variables studied. *See, e.g.*, Hanson & Chapper, *What Does Sentencing Reform Do to Criminal Appeals?*, 72 JUDICATURE 50 (1988) (three appellate courts); Martin & Prescott, *The Magnitude and Sources of Delay in Ten State Appellate Courts*, 6 JUST. SYS. J. 305 (1981). For examples at the trial level, see T. CHURCH, A. CARLSON, J. LEE & T. TAN, JUSTICE DELAYED (1978) (five to 21 courts); S. FLANDERS, CASE MANAGEMENT AND COURT MANAGEMENT IN UNITED STATES DISTRICT COURTS (1977) (six courts).

22. For example, see the variables listed in Table 1.

23. This is illustrated by the high significance of the state effects (state dummy variables) in the regressions presented in Tables 2 and 3. The state effects represent factors that differ between states but are not captured by the independent variables entered into the regression.

24. For a description of this problem, see J. MARTIN & E. PRESCOTT, APPELLATE COURT DELAY: STRUCTURAL RESPONSES TO THE PROBLEMS OF VOLUME AND DELAY 50-51 (1981).

25. *See, e.g., id.* at 42-49.

for nonpublished opinions are less complex and thus are processed more quickly.[26]

B. Acceptable Research Designs

Social science research texts consider three research designs, if properly executed, to be acceptable when studying social causation.[27] These designs are the pure experiment, the single time series, and the multiple time series. Researchers mitigate causal ambiguity either by studying the impact of an innovation over several years or by applying the innovation to a randomly selected group of courts or cases.

1. Pure random experiment— The preferred research design for causal analysis is the pure experiment, where subjects (e.g., cases or courts) are divided randomly into control and experimental groups. The "experiment" is applied to the experimental group, and the output measure (e.g., delay or number of decisions) is compared to that in the control group. The strength of this design is based on the assumptions that the two groups differ systematically in only one respect, the application of the experiment, and that all other differences are random. The Federal Judicial Center and the National Center for State Courts have been able to estimate the effectiveness of appellate settlement conferences by randomly assigning some cases to the conference procedure and other cases to a control group.[28] But successful experimental studies in courts are extremely rare because of

---

26. A similar problem is involved in studies finding that fast-track procedures, which give scheduling priority to less complex appeals, result in quicker decisions in those appeals. Such findings cannot support the conclusions that the procedures reduced delay in the courts because the speedier decisions may be at the expense of greater delay in other decisions. *See* Chapper & Hanson, *supra* note 18, at 706-70; Olson & Chapper, *supra* note 18, at 24-28.

27. *See generally* D. CAMPBELL & J. STANLEY, *supra* note 7, at 13-24; T. COOK & D. CAMPBELL, *supra* note 7; Lempert, *supra* note 7, at 127-32; Lind, Shapard & Cecil, *supra* note 7, at 91-96, 107-10.

28. J. GOLDMAN, AN EVALUATION OF THE CIVIL APPEALS MANAGEMENT PLAN: AN EXPERIMENT IN JUDICIAL ADMINISTRATION 8-11 (1977) (Federal Judicial Center publication); A. PARTRIDGE & A. LIND, A REEVALUATION OF THE CIVIL APPEALS MANAGEMENT PLAN 31-42 (1983) (Federal Judicial Center publication); D. STEELMAN & J. GOLDMAN, THE SETTLEMENT CONFERENCE: EXPERIMENTING WITH APPELLATE JUSTICE 55-58 (1986) (National Center for State Courts Publication). These studies, however, suffer from the problem that the results are peculiar to the courts studied, and judges cannot assume that settlement conferences in their courts will produce similar results. Also, the conclusions reached can depend on the particular measure of "success" used. *Compare* J. GOLDMAN, *supra*, at 36-46, *with* Goldman, *The Civil Appeals Management Plan: An Experiment in Appellate Procedural Reform*, 78 COLUM. L. REV. 1209, 1221-22 (1978).

many practical problems. Judges may be reluctant to permit researchers to dictate court operations to the extent required to ensure random assignment. The experiments require substantial sample sizes (typically several hundred, depending on the magnitude of the impact found), so a true experiment may require several years.

By far, the most important problem with pure experiments in court research is that the experimental and control groups may be treated differently in ways other than the "experimental" treatment, thus contaminating the results. Researchers can use the experimental design to determine whether appellate settlement conferences foster more settlements because the conferences take place in the early stages of case processing, and the settlement decisions are made by *attorneys,* not by the court. If aspects of court processing are controlled by judges or court staff, a true experiment cannot be used to evaluate them because the researcher cannot determine whether the judges or court staff decide independently of the "experimental" procedure.[29] For example, studying the impact of staff attorneys by randomly assigning some cases to staff and others to a control group would not be helpful because judges may give staff cases priority or otherwise treat these cases differently from the control cases. It would be unrealistic to expect judges to ignore the existence of the experiment. This problem could be avoided if the court, rather than the case, is the unit of analysis, but that would require an enormous research effort, using hundreds of courts with varying procedures randomly assigned to them.

2. *Long time series*— The other generally accepted research designs for the social science exploration of causation are based on time differences between variables, and they rely heavily on econometric techniques. A change in court operations in one year can cause, but cannot be caused by, changes in the dependent variable—in the present research, decision output—in subsequent years. The time-series design, for example, could study whether output increases significantly in the years after staff attorneys were hired. This design differs from the discredited before-and-after design mainly because the data are obtained for enough time periods to help control for the impact of factors other than the evaluated change. The rule of thumb is that

29. For example, the experimental design cannot be used to determine whether settlement conferences reduce delay because other procedural hurdles and/or the priorities involved in scheduling conference cases may affect the scheduling of the nonconference cases. *See* A. PARTRIDGE & A. LIND, *supra* note 28, at 49-59; D. STEELMAN & J. GOLDMAN, *supra* note 28, at 55-91.

time-series tests require data from at least fifty time periods.[30] Such large data sets are rarely available for court studies.[31]

3. Multiple time-series design— The final acceptable design, and the design used here, is the multiple time-series analysis, which uses data from many courts over a period of several years. This has long been considered among the best social science research designs.[32] Nevertheless, the design is seldom used for court studies, probably because it is more the province of economists than of the disciplines whose members generally study courts.

In the present study, the application of the multiple time-series design involves data for forty-four state appellate court systems over a time span of seventeen years for most. The variables include measures of decision output and of the changes being evaluated. In this way, the time series for the different court systems, each of which is not long enough for proper analysis, can be combined into a regression with a large sample size.[33]

The only multiple time-series study of appellate court decision output—in fact, the only prior study of state appellate procedures other than settlement conferences that uses an acceptable research design—was conducted more than half a century ago. In this excellent, large-scale project at the University of Michigan Law School, the authors counted opinions published in most states during the first three decades of the century and evaluated efficiency measures, such as adding lawyer staff and deciding cases without opinion, by determining whether opinion output increased in the years after a particular measure was adopted by the various courts.[34] The authors of that work, however, did not have the statistical techniques or computing power to control for the impact of other changes. The present study attempts to modernize that seminal project.

---

30. T. COOK & D. CAMPBELL, *supra* note 7, at 228.

31. For an example of a study with sufficient time series data, see Garner, *Delay Reduction in the Federal Courts: Rule 50(b) and the Federal Speedy Trial Act of 1974*, 3 J. QUANTITATIVE CRIMINOLOGY 229 (1987).

32. Campbell and Stanley, two highly regarded methodologists, consider this "an excellent quasi-experimental design, perhaps the best of the more feasible designs." D. CAMPBELL & J. STANLEY, *supra* note 7, at 57; *see also* T. COOK & D. CAMPBELL, *supra* note 7, at 214-18. Lempert considers this design to be the "design *par excellence*" for impact research. Lempert, *supra* note 7, at 130.

33. The statistical methodology used is further described in Part II.

34. Curran & Sunderland, *The Organization and Operation of Courts of Review: An Examination of the Various Methods Employed to Increase the Operating Capacity and Efficiency of Appellate Courts*, in THIRD REPORT OF THE JUDICIAL COUNCIL OF MICHIGAN 52 (1933).

II. RESEARCH PROCEDURES

This Section describes the research procedures applied to the multiple time-series model. The section pays particular attention to the statistical methods and the construction of variables.[35]

A. Statistical Methods

The present research uses a production function model and a fixed effects regression. Both are basic, routine statistical procedures that have been used and refined for several decades.

The production function has long been considered to be the best procedure for analyzing the volume of output in an organization. The production function gives the amount of output that can be produced by a given combination of labor and other inputs.[36] It assumes that output increases at a decreasing rate: that each additional unit of labor, for example, will cause the output to increase, but the rate of increase will get smaller unless all other inputs increase proportionately. For example, applying this notion to appellate courts, one would not expect that simply adding more judges would result in a corresponding increase in appeals decided, because the pool of cases to be decided, the court's staff, and other inputs have not increased. Without a production function model, the researcher would make the unrealistic assumption that, for example, ten percent more judges will decide ten percent more cases, even if no more cases were filed or no more staff attorneys or law clerks were hired.

For the regression analysis, we have used the "fixed-effects" model, which includes dummy variables for each state (dummy variables are either zero or one; here, the dummy variable for a

---

35. For more detail concerning the analyses and data used, see T. Marvell & C. Moody, State Appellate Court Adaptation to Caseload Growth (July 1986) (unpublished report; National Criminal Justice Reference Service, Washington, D.C.); *Responses to Caseload Growth*, *supra* note 3. The full data set is available from the authors or from the Inter-University Consortium for Political and Social Research, Ann Arbor, Michigan.

36. The research uses the most common form of production function, the Cobb-Douglas function, which has been widely used for half a century. For more detailed descriptions of the Cobb-Douglas function, see M. INTRILIGATOR, ECONOMETRIC MODELS, TECHNIQUES, AND APPLICATIONS 262-80, 288-92 (1978); A. WALTERS, AN INTRODUCTION TO ECONOMETRICS 269-339 (1970). For the specific application of the Cobb-Douglas function to the present research (including the mathematical functions) and more detail concerning the analysis and data used, see T. Marvell & C. Moody, *supra* note 35, at ch. 2, 1-8.

state is one for observations pertaining to the state, and zero otherwise). This, again, is standard statistical procedure.[37]

B. Unit of Analysis

State appellate court operations and productivity are best analyzed at the level of the entire appellate system, rather than at the level of an individual court, because state appellate structures vary.[38] In 1968, the initial year of the research, only seventeen of the forty-four states studied had intermediate appellate courts (IAC's). By 1984, they existed in thirty-one of the states. IAC jurisdiction differs greatly. In some states, IAC's receive virtually all initial appeals; in others, a sizeable portion of the appeals are routed directly to the state supreme court. IAC's in the former states, of course, are usually much larger than those in the latter states. A few states, moreover, have two IAC's with different subject-matter jurisdiction. Finally, the division of appellate jurisdiction changed substantially during the period of

---

37. *See* R. PINDYCK & D. RUBINFELD, ECONOMETRIC MODELS AND ECONOMIC FORECASTS 252-61 (1981); Moody & Marvell, *Appellate and Trial Court Caseload Growth: A Pooled Time-Series-Cross-Section Analysis*, 3 J. QUANTITATIVE CRIMINOLOGY 143, 150-52 (1987) [hereinafter *Cross-Section Analysis*]. The multiple time-series design, it should be noted, is also called the time-series cross-section design, and the fixed-effects model is also called the "least squares dummy variable" or the "covariance" model. The fixed-effects model also includes separate dummies for each year; these were eliminated because, as a group, they were not significant. *See* R. PINDYCK & D. RUBINFELD, *supra*, at 255. This means that it is very likely that there are nationwide factors affecting appellate decision output that are not represented by the independent variables entered in the analysis.

As a time-series analysis, this research occasionally encountered autocorrelation problems that were corrected using the normal procedure. R. PINDYCK & D. RUBINFELD, *supra*, at 258-61. The regression in Table 2 and the supreme court analysis in Table 3 had Durbin-Watson statistics of 1.32 and 1.66, respectively, before correction (values below 1.78 indicate possible autocorrelation problems). The Breusch-Pagan test uncovered a slight heteroskedasticity problem (the test looked at whether variance is related to court size and found variance is greater in small states), which was eliminated by weighting the regression by the square root of population. Tests for multicollinearity (independent variables highly correlated) revealed a possible problem with respect to the number of filings for the current year and the prior year. That is, the F-Ratios for these variables given in Tables 2 and 3 may be too low, and the apportionment of the coefficient between the two may be slightly off. The only other multicollinearity is between state dummies (and between the dummies and the intercept), which is not relevant to the results here. Application of the Granger-Sims test found no simultaneity (independent variable causing independent variables), except for a possible problem with respect to the number of judges, discussed *infra* note 58. For a more complete and technical analysis of the application of the fixed-effects model to the present research, see T. Marvell & C. Moody, *supra* note 35, at ch. 2, 4-9.

38. The variations are described in Marvell, *supra* note 5, at 90-97.

this study; most states have created or greatly expanded IAC's since 1968.

These changes render trends in individual courts misleading because the content of their caseloads change. When a state transfers jurisdiction to an IAC from a supreme court, the supreme court typically retains the more difficult appeals and the more time-consuming lawmaking function of appellate courts, while the IAC concentrates on error correction. Hence, the caseload and procedure of a supreme court often cannot be compared to those of earlier years or to those of supreme courts in other states. IAC operations are more comparable because all IAC's concentrate on error correction, but states differ concerning how many of the difficult error-correcting cases are retained in the supreme court.

By focusing on the whole appellate system, one can avoid the troubles caused by shifting jurisdiction.[39] Taken at the appellate level, the variables are derived by adding figures from the supreme courts and the IAC's in each state. For example, the number of decisions per judge is the sum of supreme court and IAC decisions divided by the sum of supreme court and IAC judges, and the percentage of opinions that are unpublished is the total number of unpublished opinions in the two court levels divided by the total number of opinions.

As discussed later, the study also analyzed separately three different categories of appellate systems: (1) supreme courts in states without IAC's; (2) appellate systems that have large IAC's—those taking at least eighty percent of the appeals in the state; and (3) mid-range states, with IAC's deciding less than eighty percent of all appeals. Variable means presented in Table 1 show that there are some major differences among these three types of appellate systems.[40]

---

39. *See infra* note 54 concerning controls when there are changes in the jurisdiction of the overall state appellate system, as opposed to changes in the jurisdiction of individual courts.

40. *See infra* text accompanying note 47. Table 1, however, exaggerates the differences. Because many states created IAC's, data for supreme courts alone tend to be for earlier years in the study, when decisions per judge were lower and efficiency measures were less drastic. The same applies to the relationship between mid-range appellate systems and states with large IAC's.

C. *The Dependent Variable: Number of Appeals Decided Per Judge*

The two major types of variables in a regression analysis are the dependent and independent variables. Each regression has one dependent variable, placed on the left side of the equation. The independent variables, on the right side of the equation, are assumed to "cause" changes in the dependent variable.[41] The dependent variable in this study is the number of appeals decided per judge during a year, and the following paragraphs define the two elements of this term.

1. Decisions— Appeals "decided" are those resolved on the merits—with or without written opinion.[42] "Decisions" do not include denials of writs or certiorari petitions, appeals withdrawn or dismissed for lack of progress, determinations made on procedural motions, or rehearing petitions. "Decisions" do include, however, cases dismissed by written opinion, disciplinary cases given full review, appeals summarily decided on the merits (e.g., cases resolved without briefs), and certiorari petitions or other discretionary writs granted and then considered in the manner of ordinary appeals.[43]

---

41. For detailed information about the variables and the sources of data, see T. MARVELL & C. MOODY, *supra* note 35, at chs. 6-12.

The major sources of data are state court administrative offices, unpublished reports prepared by the courts, and the large body of secondary literature on the topic, *see supra* note 5. Especially useful were several nationwide surveys of appellate court operations, including W. KRAMER, COMPARATIVE OUTLINE OF BASIC APPELLATE COURT STRUCTURE AND PROCEDURES IN THE UNITED STATES (1983), and Council of Chief Judges of Cts. of Appeals, Chief Judges as Administrators: A Survey (Apr. 1984) (unpublished report) (Appellate Judges Conference of the Am. Bar Ass'n). In several states, where all opinions are published, decision data were obtained by counting opinions. Court clerks were interviewed in all states studied, as well as staff attorneys or judges in most states. The completed data were sent to the clerks for comments.

42. The number of decisions for most courts is the number of cases decided, but, for some, it is the number of opinions or orders deciding cases. The latter measure excludes cases consolidated for decision in one opinion (which amount to approximately five to ten percent of all appeals decided). The decision measure never varies within a state; thus, in the regression analysis, the state dummy variables control for the difference in methods of counting decisions.

43. The composition of appeals probably varies greatly between states; thus, the difficulty of appeals may vary, and the number of appeals decided may not be comparable between states. The regression analysis, however, controls for such differences by entering state dummy variables. There is no reason to believe that the types of appeals in state courts have changed appreciably during the relatively short period of this study. If there were nationwide changes, they would be controlled for by the year dummies (which would then become significant). The nationwide portion of criminal appeals, which probably take less time than civil appeals, has changed little since the mid-1960's, but the portion of criminal appeals did vary in individual states. *See* BUREAU OF JUST. STATISTICS,

2. *Judges*— The "number of judges" is the number of sitting, full-time appellate judges. It excludes vacancies but includes judges temporarily absent due to illness or recusal. It excludes senior or retired judges, unless they sit full-time and are regular judges for all practical purposes.[44]

III. RESULTS

The analysis evaluates the effectiveness of changes that courts have adopted to increase decision output, emphasizing those that the literature has recommended as effective answers to caseload problems.[45] In addition, the analysis includes several control variables to represent changes, other than efficiency measures, that may affect output. In all, the analysis contains nineteen independent variables (not including the state effects) that are organized into eight categories below.[46] The variable

---

supra note 1, at 2. We have information concerning the number of criminal cases for most states in the analysis; when the percentage of criminal cases was entered in the regression analysis, the variable was far from significant. As discussed *infra* note 54, the analysis controls for changes in caseload compensation when appellate jurisdiction changes.

44. The major sources of information about judges were state court annual reports, state judicial directories and biographies, and the lists of judges printed in West Publishing Company state reporters.

45. For examples of the enormous body of literature available on this topic, see *supra* note 5. The analysis includes virtually all major changes appellate courts have made to increase output (excluding technological changes such as word-processing equipment). One exception is the Washington Court of Appeals motion on the merits procedure, under which staff attorneys make initial decisions in many cases, with further review by the judges available. With rare exceptions, cases were not decided under this program until 1985. *See* Green & Keyes, *Motion on the Merits: An Effective Response to Appellate Congestion and Delay*, 70 JUDICATURE 168 (1986). Another possible exception is the use of specialized panels for criminal cases (in New Jersey during 1985) or for workers' compensation cases (Arizona since 1969 and Illinois since 1985). Only one of these changes occurred during the time period studied here. Another procedure designed to help appellate courts deal with rising caseloads is the preargument settlement conference; this was not included in the study because the primary aim of such conferences is to reduce the volume of cases submitted to the judges, not to increase decision output. "Screening" and "fast-track" procedures are not included as variables because they do not involve efficiency measures distinct from the variables that are included; rather, they are procedures for determining which cases are subjected to the various efficiency measures. Screening, as the name implies, involves decisions by the court as to which cases are argued, decided with opinion, or the like. Fast-track procedures apply certain efficiency measures in return for scheduling priority.

46. For background information, see *Responses to Caseload Growth*, *supra* note 3. For descriptions of the extent of changes made by each state studied, see T. MARVELL & C. MOODY, *supra* note 35, at Pt. IV, and *Responses to Caseload Growth*, *supra* note 3. This Article does not fully describe the numerous arguments as to why these various

means are given in Table 1.[47] Table 2 presents the results of the estimation for the full sample, and Table 3 presents the results for three separate categories of states.[48] The three are discussed separately because statistical tests revealed significant differences among the production functions for the different categories of states.[49] However, because the differences are not major and do not affect most variables, discussion will concentrate on full analysis in Table 2, which has more observations and, therefore, more precise results.

---

changes may or may not add to court productivity. For sources discussing this point, see *supra* note 5.

47. As indicated in Table 2, several variables were entered in the analysis in logged form. The mean of the (natural) log of these variables (for all states, states without IAC's, mid-range states, and states with large IAC's) is: decisions per judge (4.07, 3.79, 4.06, 4.50); filings per judge (4.42, 4.19, 4.34, 4.82); number of judges (2.48, 1.71, 2.83, 3.32); staff attorneys per judge (.21, .15, .16, .42); law clerks per judge (.79, .76, .76, .87). For the number of staff attorneys and law clerks per judge, one is added to the variables before taking the log because it is impossible to take the log of zero, and adding one sets the lower limit of the logged variable at zero.

48. *See supra* text accompanying note 40. The 80% level was selected because it is a natural break: States did not move through that level without jurisdiction changes. These regressions, as we previously explained, encompass whole appellate systems. The results of the regression for states with large IAC's are nearly the same as the results for the IAC's alone (that is, without including the supreme court judges, decisions, and so on). A similar comparison for mid-range states is not feasible because the jurisdictions of the supreme courts and IAC's shift widely and because, in some states, total filings, the most important variable, cannot be meaningfully apportioned to either court level.

49. The Chow test was used to determine whether the results for each of the three types of appellate systems are significantly different from the remainder of the courts. *See* R. PINDYCK & D. RUBINFELD, *supra* note 37, at 123-26. Under the Chow procedure the analysis represented in Table 2 was conducted three times, each testing 1) whether the results for states with one type of appellate system are significantly different from the results for the remaining states with respect to each of the variables; and 2) whether the coefficients as a group are significantly different (excluding the IAC percentage variable, which defines the groups of states). The latter test showed significant differences, although the differences were not greatly significant, for all three types of court systems. The F-Ratios are 2.62 for supreme courts in states without IAC's, 1.96 for mid-range appellate systems, and 1.83 for states with large IAC's.

With respect to individual variables, significant differences (to the .05 level) were found for the number of judges (mid-range and large IAC's), extra judges (supreme courts without IAC's and mid-range systems), intermediate court percent (mid-range and large IAC's), panel size (mid-range IAC's), argument length (supreme courts), and law clerks (supreme courts).

Table 1. Variable Means

Independent Variable	All States	States without IAC	Mid-range States	States with large IACs
decisions per judge	58.56	44.26	57.97	90.02
filings per judge	83.10	66.02	76.71	123.97
prior year filings per judge	76.71	61.56	70.11	115.58
number of judges	11.94	5.53	16.95	27.66
extra judges (dummy)	.26	.23	.19	.40
percent new judges	11.45	9.57	14.18	10.60
staff attorneys per judge	.23	.16	.17	.52
law clerks per judge	1.20	1.14	1.14	1.39
percent unpublished opinions	22.04	9.35	22.89	39.94
15% memo opinions (dummy)	.44	.36	.41	.64
50% memo opinions (dummy)	.15	.10	.15	.30
percent decided without opinions	8.58	9.41	4.74	11.06
intermediate court percent	42.61	0	61.46	90.57
average panel size	4.30	4.86	4.15	3.53
15% w/o oral argument (dummy)	.59	.42	.65	.79
50% w/o oral argument (dummy)	.19	.10	.30	.19
oral argument length (minutes)	59.58	64.23	58.22	55.16
10% summary decisions (dummy)	.07	.12	.07	.02
jurisdiction change (dummy)	.05	0	.09	.05
change in procedure for docketing filings	1.92	1.83	2.27	1.74

The regression includes forty-four states, with five to seventeen years for each.[50] Missing data for some independent variables, especially filings, preclude the use of all observations with data for decision output. Still, all but four states have complete data for at least ten years in the main analysis (Table 2). In the analysis of the three appellate system types (Table 3), nearly half the states have time series of under ten years because jurisdiction changes caused many states to be shifted from one category to another.[51] The results are robust in that they do not change substantially when using different variable specifications or

50. The District of Columbia is counted as a state. The states not included in the analysis are Indiana, Minnesota, Nevada, Ohio, Oklahoma, Tennessee, and West Virginia. Decision data from 1968 to 1984 were obtained for 38 of the 44 states, and decision data back to at least 1972 were obtained for the remaining six states.

51. States were deleted from these analyses if they had less than five years of data.

regression models.[52] It is important to note that the state dummies in all regressions are highly significant, which suggests that productivity differs greatly between state appellate systems (but not over time within states[53]) for reasons not represented by other variables in the analysis. One possible reason might be a disparity in how hard judges work.

The following sections briefly describe the independent variables[54]

---

52. The results for different types of analysis are given in T. Marvell & C. Moody, *supra* note 35, at chs. 14-15. That study, however, differs from the present one in that it does not include four states added afterwards, does not distinguish between staff attorneys and law clerks, does not include current-year filings, and does not correct for heteroskedasticity. Of particular importance is the fact that the results are very similar when using a linear regression, without using logged variables. Although this procedure is more intuitively appealing to the layperson, it is not the best procedure for reasons given in Part IIA. The results of this linear regression are:

Independent Variable	Coefficient	T-ratio
filings per judge	.086	3.07
prior year filing per judge	.381	12.41
extra judges (dummy)	4.004	2.23
percent new judges	-.016	.70
staff attorneys per judge	3.771	1.14
law clerks per judge	1.594	.93
percent unpublished opinions	.181	3.81
15% memo opinions (dummy)	.571	.39
50% memo opinions (dummy)	3.095	1.58
percent decided without opinions	.327	4.92
intermediate court percent	.246	5.78
average panel size	1.004	1.22
15% w/o oral argument (dummy)	.004	.00
50% oral argument (dummy)	6.550	3.92
oral argument length	.094	1.51
10% summary decisions (dummy)	.634	.24
jurisdiction change (dummy)	-2.630	.82
change in procedure for docketing filings	.083	3.76

Number of observations, 541; degrees of freedom, 479; F-Ratio, 791; Durbin-Watson statistic, 2.08 (1.64 before correction); F-Ratio for state dummies, 32.00.

53. See *supra* note 37 for a discussion of year dummies.

54. Besides the variables of interest, the regression includes several "nuisance" variables that control for jurisdiction changes and changes in procedures for counting appeals. During the period of the study, new laws in six states provided for direct appeal to

and discuss the results presented in Tables 2 and 3.[55]

1. Filings— The "raw material" input is measured by two variables, the number of appeals filed per judge for the current year and for the prior year. This input includes a lagged variable because the typical court takes approximately a year to decide cases. The current year variable is included because a larger backlog of pending cases may prompt judges to expand output. Filings are appeals of right and, with rare exceptions, do not include writs and certiorari petitions, which are ordinarily dismissed without decision on the merits.[56]

The results show a very strong relationship between filings, especially the previous year's filings, and decision output. Ten percent more filings will lead to approximately six percent more decisions,[57] even if the court makes none of the other changes studied. In other words, courts hand down more decisions simply because more cases are filed. Similar results occur in all three types of court systems.

---

appellate courts from limited jurisdiction courts (previously, these appeals went to general jurisdiction trial courts). The new laws generally cover criminal misdemeanors and small civil cases. Because these cases are probably less complex than most appeals, their addition can cause total productivity to rise. The changes occurred when intermediate appellate courts were created or expanded. Thus, analysis without this control variable would overstate the impact of IAC's on productivity.

The second nuisance variable controls for changes in methods of counting appeals filed. Some courts count appeals when the notice of appeal is filed, while others count them when the record arrives. The latter system reduces the volume of appeals because many cases are dropped between notice of appeal and record completion. Six of the states studied adopted the first counting system during the study period, artificially increasing filings. To control for this, the analysis includes an interaction variable—filings per judge logged times a dummy variable indicating the type of docketing (0 = when the record arrives; 1 = at notice of appeal).

55. Readers not familiar with statistical analysis should note that the findings are, to varying degrees, imprecise. It is likely that the impact of some changes varies between courts, and the results here estimate the average impact. In general, a finding that a change shows no statistically significant impact on decision output indicates that it is unlikely that the change has more than slight impact in any courts, although this conclusion is less clear when the degrees of freedom are less (e.g., in Table 3). When a relationship is significant, the coefficient indicates the midpoint of a range in which the actual results are likely to fall. This range is larger when the T-Ratio is smaller. For example, the range (95% confidence interval) for law clerks in Table 1 is between .009 and .207 (or plus or minus 92% of the .108 coefficient), indicating that the findings are very imprecise. In contrast, the range for prior year filings per judge is .363 to .525, a spread of plus or minus 18%. The differences between results in Table 2 and Table 3 are due largely to the general impreciseness of the findings, especially those in Table 3. It would be risky to attribute the different results for different court systems (in Table 3) to actual differences between courts unless there are strong reasons for believing that such relationships exist.

56. The filing data were obtained from a Bureau of Justice Statistics study. *See* BUREAU OF JUST. STATISTICS, *supra* note 1, at 4-6.

57. In statistical terms, the "elasticity" is .6. Elasticity is the same as the coefficient when both variables are logged. More exactly, the elasticity for filings is the sum of the value for the current and prior year (.180 + .444 = .624).

2. *Judges*— Three variables pertain to judges. First, the number of appellate judges is a variable that indicates returns to scale—that is, whether adding more judges produces a corresponding increase in output. If the coefficient is significantly different from zero, then the returns to scale are significantly different from unitary. Taken overall, as shown in Table 2, there are no positive or negative returns to scale,[58] but the result differs among court systems (Table 3). For states without IAC's, there are decreasing returns to scale, although at a low significance level. Here, a one hundred percent increase in supreme court judges, legal staff, and filings will probably result in less than a one hundred percent increase in output. This result makes intuitive sense: the need for all judges to participate in important decisions at the top level poses greater problems than found in other courts.[59]

Second, the number-of-extra-judges variable refers to the temporary assignment of trial judges or retired judges to supplement regular judicial capacity.[60] Courts in most states used this practice at one time or another, but the courts seldom used extra judges to increase the effective number of judges by more than ten percent. Because statistics on the precise number of extra judges are not always available, a dummy variable[61] signifies whether extra judges were used.

---

58. Applications of the Granger-Sims test found a possible simultaneity (reverse causation) problem with respect to the number of judges. That is, if a state appellate system produces more (less) decisions per judge in a particular year, the state government may be more (less) reluctant to increase the number of judges that year. For a general discussion of this type of problem, see Granger, *Investigating Causal Relations by Econometric Models and Cross-Spectral Methods*, 37 ECONOMETRICA 424 (1969). In practice, such a relationship is unlikely. First, the simultaneity effect is not highly significant (T-Ratio = 2.30, just above 1.96, which indicates significance to the .95 level), and with 16 tests for the different independent variables, one would expect that, just by chance, one would produce significant results. Second, common sense suggests that the causal relationship is improbable. Legislatures may well be less likely to create new appellate judgeships if the courts are able to handle the current caseload by increasing productivity, but the impact would be lagged rather than simultaneous. There is considerable time between a change in productivity and any impact on the number of judges, because it takes a while for legislatures to react to the changes, for the law to reach its effective date, and for new judges to be appointed or elected.

59. *See* Dethmers, *Delay in State Courts of Last Resort*, 328 ANNALS 153, 158 (1960).

60. This should be distinguished from the more common practice of using trial and retired judges to fill in when regular judges are temporarily unavailable.

61. Dummy variables are either zero or one. In this situation, the variable is zero if the appellate system does not use extra judges in a particular year, and one if it does. The coefficient of the dummy variable is determined solely by within-state changes. For the extra-judge dummy variable, there were 18 such changes in the analysis for Table 2, and, in Table 3 there were 10 changes for states without IAC's and seven for mid-range states. There was only one change for states with large IAC's, so the results are particularly uncertain. Of the 44 states in the analysis, 41 have data on the judge equivalents added by using extra judges. In an analysis (comparable to that in Table 2) with just these states, the coefficient (elasticity) of the extra judge variable is .66 (T-Ratio = 3.34). Because the coefficient is less than one, this suggests that extra judges do not contribute as much as regular judges to decision output for a given amount of judge time.

Table 2. Regression Analysis of All States

Dependent Variable: decisions per judge (logged)

Independent Variable	Coefficient	T-ratio
filings per judge (logged)	.180[a]	4.38
prior year filings per judge (logged)	.444[a]	10.87
judges (logged)	-.018	.37
extra judges (dummy)	.059[c]	2.50
percent new judges	-.0007[c]	2.39
staff attorneys per judge (logged)	-.001	.01
law clerks per judge (logged)	.108[c]	2.15
percent unpublished opinions	.0026[a]	4.18
15% memo opinions (dummy)	.016	.82
50% memo opinions (dummy)	.045[d]	1.77
percent without opinions	.0040[a]	4.87
intermediate court percent	.0028[a]	4.10
average panel size	.005	.43
15% w/o oral argument (dummy)	.040[d]	1.82
50% w/o oral argument (dummy)	.036	1.55
oral argument length	.001	.91
10% summary decisions (dummy)	.000	.01
juris. change (dummy)	.042	.95
change in procedure for docketing filings	.015[c]	2.02
Observations	541	
Degrees of freedom	478	
F Ratio	2838	
DW	2.04	
F Ratio for state dummies	16.1	

Notes: a - significant to the .0001 level; b - to the .01 level; c - to the .05 level; d - to the .10 level.

Table 3. Regression Analysis, Different Types of Appellate Systems

Dependent Variable: decisions per judge (logged)

Independent Variable	Coefficients (and T-Ratios)					
	Supreme Courts (states without IACs)		Mid-Range Appellate Systems		States with Large IACs	
filings per judge (logged)	.148[c]	(2.06)	.315[a]	(4.14)	.213[b]	(2.70)
prior year filings per judge (logged)	.429[a]	(6.09)	.425[a]	(4.91)	.426[a]	(4.92)
judges (logged)	−.211[d]	(1.71)	−.076	(.74)	−.056	(.70)
extra judges (dummy)	.146[b]	(3.71)	.005	(.15)	.063	(1.47)
percent new judges	−.0008	(1.14)	−.0009[d]	(1.66)	−.0003	(.51)
staff attorneys per judge (logged)	−.036	(.40)	.135	(.94)	.244[c]	(2.15)
law clerks per judge (logged)	.053	(.75)	.222[c]	(2.29)	.173	(1.34)
percent unpublished opinions	.0071[a]	(4.52)	.0002	(.19)	.0012	(1.18)
15% memo opinions (dummy)	.015	(.46)	−.024	(.51)	.009	(.30)
50% memo opinions (dummy)	−.035	(.43)	.038	(1.12)	.058	(1.42)
percent decided without opinions	.0033[d]	(1.85)	.0063b	(3.00)	.0017	(1.38)
intermediate court percent	—	—	.0015	(1.32)	.0093[c]	(2.51)
average panel size	−.026	(.54)	.016	(1.26)	.053[c]	(2.19)
15% w/o oral argument (dummy)	.121[b]	(3.04)	−.014	(.36)	.020	(.57)
50% w/o oral argument (dummy)	.003	(.07)	.240[b]	(2.71)	.027	(.85)
oral argument length	−.001	(1.21)	.000	(.38)	.003	(.73)
10% sumary dec. (dummy)	−.019	(.42)	−.236[c]	(2.15)	.106[c]	(2.21)
juris. change (dummy)	—	—	−.026	(.35)	−.157[c]	(2.56)
change docketing filings	.035[c]	(1.99)	−.141[d]	(1.76)	.016[d]	(1.92)
Number of observations	220		151		154	
Degrees of freedom	184		115		121	
F Ratio	1419		98		221	
Durbin Watson statistic	1.96		2.09		1.80	
F Ratio for state dummies	10.87		8.42		13.34	

Notes: a − significant to the .0001 level; b − to the .01 level; c − to the .05 level; d − to the .10 level.

The use of extra judges is significant in the overall analysis (Table 2). In the disaggregated analyses, however, it is significant only in states without IAC's, possibly because in these states, the use of temporary judges provides a greater proportional addition to existing judicial resources.

Third, the rate of judge turnover is expressed by the percentage of judges who are new to the court in the year studied.[62] This variable was included because new judges may be less productive than experienced judges. The percent-new-judges variable has a significant negative impact on productivity, as expected, but the magnitude is very small.

3. Legal staff— The legal staff consists of staff attorneys and the judges' personal law clerks. The former work for the whole court, and the latter work for individual judges.[63] Here, as elsewhere, the variables are for the whole appellate system; law clerks per judge, for example, are the total number of appellate law clerks in the state divided by the total number of judges on the supreme court and IAC (Table 2).

Adding more staff attorneys and law clerks has surprisingly little impact on decisions. For law clerks, the impact is significant in the overall analysis, but the magnitude is small. The coefficient of .108 means that if a second law clerk is given to each judge, the decision output per judge would increase by only some ten percent.[64] The nationwide analysis in Table 2 found that staff attorneys have no impact on decision output. Table 3, however, suggests that staff attorneys may increase productivity in large IAC's. Because the significance level is low, this is far from certain, but it is quite possible that judges in large IAC's are more likely to delegate decision tasks to the staff.

Although the data provide no explanation for the meager impact of legal staff on productivity, one can speculate that the

---

62. More precisely, it is the number of new judges in the last month of the prior year and the first 11 months of the current year, divided by the total number of judges. The variable is lagged a month because it usually takes at least one month for the work of a new judge to appear as decision output.

63. Information about attorney aides, more than other variables, came from secondary sources and from interviews with court personnel. In several states, court staff themselves obtained the data from budget statements for the study.

64. This is a very imprecise estimate because the significance level is low. The analysis without logged variables found that law clerks did not have a significant impact. *See supra* note 52. A similar result was recently obtained in a regression analysis of United States Supreme Court productivity; the authors found that adding law clerks had a slight *negative* impact on opinion output in the period 1948-1985, which they attributed to the administrative costs of managing a larger staff. Stewart & Heck, *Caseloads and Controversies: A Different Perspective on the "Overburdened" U.S. Supreme Court*, 12 JUST. SYS. J. 370, 378-379 (1987).

tasks performed by the staff often add to the work of the judges. For example, a major responsibility of the staff is to gather legal authorities and other information,[65] thus adding to the amount of material for the judges to consider. If this is true, attorney staff increase the quality, more than the quantity, of decisions. The present research, of course, does not address issues of quality.

4. Opinion practices— The analysis includes variables pertaining to three types of opinion practices, all of which can increase productivity because they reduce the amount of time that judges spend writing opinions. The first is the percentage of cases decided with unpublished opinions—that is, with opinions written but not published in the official reports or in the West regional reports.[66] This is an effective efficiency measure. Reducing opinion publication has a highly significant impact, although possibly greater in states without IAC's.

The second practice is issuing opinions that are not signed by the authoring judge. These are typically called per curiam, memorandum, or simply "memo" opinions—the latter term is used here to refer to all unsigned opinions. The use of memo opinions represents a crude measure of opinion length. Most are much shorter than regular signed opinions, but several courts issue lengthy memo opinions, and a few courts issue very short signed opinions. Because some courts do not issue statistics on the actual number of memo opinions, their use is expressed in dummy variables: One signifies whether memo opinions constitute fifteen percent or more of the total number of opinions, and a second signifies whether they constitute at least half the opinions.[67] The analysis found that issuing memo opinions has only a small impact, slightly significant for one of the variables in Table 2, but never significant in Table 3.[68] However, the analysis might

---

65. T. MARVELL, APPELLATE COURTS AND LAWYERS: INFORMATION GATHERING IN THE ADVERSARY SYSTEM 88-97 (1978).

66. Information for this variable was usually available from court statistics, and it was checked against figures supplied by West Publishing Company for the number of opinions it published for each court in each year. For several states, the West figures were the only information available concerning the extent of publication.

67. Dummy variables here and elsewhere were "smoothed" in the sense that they were not changed if the court only temporarily and slightly passed the dividing line.

68. There were 30 within-state changes in the 15% dummy variable in Table 2, and 13 for the 50% dummy variable. Comparable figures for Table 3 are 11 and three for supreme courts without IAC's, 10 and five for mid-range states, and five and three for states with large IAC's. Some changes are not included in Table 3 because they occurred when appellate systems changed categories. Information about the number of memo opinions was obtained for 37 of the 44 states. In an analysis with only these states, the

have shown a greater impact if a more refined measure of opinion length were available.

The final opinion variable is the percentage of appeals decided without any opinion. These are decisions on the merits (denials of discretionary appeals and writs are excluded in the measure of decisions) that are decided by a simple order giving only the bare holding. Deciding cases without opinion greatly increases productivity (Table 2). The results in Table 3 suggest that the most pronounced impact may occur in mid-range appellate systems, although there is no ready explanation for why this might be so.

5. Intermediate courts— Because IAC jurisdiction—and therefore the portion of cases decided by IAC's and supreme courts—varies greatly, IAC's are measured by the percentage of appellate system decisions made there. The percentage, of course, is zero in states without IAC's, and it varies from fifty to ninety-five percent for nearly all states with IAC's. The IAC variable is significant for the overall regression (Table 2).[69] This means that increasing IAC jurisdiction increases the overall productivity of state appellate systems even after controlling for all the other changes studied here.

6. Panel decisions— With few exceptions, IAC's either have only three judges or they decide cases in panels of three judges. Most supreme courts sat in panels of three to five judges at some time during the study. One would expect that the involvement of fewer judges in decisions would increase productivity. In the analysis, the average size of panels is the mean number of judges deciding cases. If a court decides all cases en banc (i.e., without using panels), the panel size is considered to be the total number of judges on the court. The average panel size for states with IAC's is computed from the panel sizes for the supreme court and IAC, weighted for the relative decision output of each.

Perhaps the most surprising finding in this research is that reducing panel size does not increase productivity.[70] A possible explanation is that the judge assigned to write the opinion does most of the work on an appeal, and smaller panels do not reduce

---

percent of cases decided with memo opinions was significantly related to decision output (coefficient = .0016, T-Ratio = 2.03).

69. This result cannot be compared to those in Table 3 because the extent of IAC involvement in producing appellate decisions is the criterion for distinguishing the three types of appellate systems. The results in Table 3, therefore, do not reflect the impact of major changes in IAC use.

70. These results, like all the findings here, apply only within the range experienced by the courts studied. The average panel size is almost always limited to the three to

opinion assignments or, probably, the amount of time required to write opinions.[71] Furthermore, each concurring judge may scrutinize the opinion more carefully when the panel is small, knowing that there are fewer judges who can do so.

7. *Oral arguments*— Oral arguments can be curtailed by reducing their frequency or their length. The available information permits only rather crude variables for both. The number of arguments is represented by two dummy variables signifying whether at least fifteen percent and fifty percent, respectively, of the cases are decided without argument. The length of arguments is represented by the time limit specified in court rules, but this only approximates the actual time spent because many attorneys do not use their allotted time.

Reducing the percentage of cases argued produces a significant impact on productivity,[72] although the results differ somewhat between appellate systems (Table 3). The oral argument time limits have no discernible impact on productivity.

8. *Summary judgment procedures*— Summary judgment procedures, the most radical answer to caseload growth, are similar to such procedures at the trial level. The court or one party moves to affirm before the parties file their briefs (in some courts, even before the transcript is filed), and the information received by the court is usually limited to short memoranda from the parties. When the appeal is summarily decided, therefore, the court dispenses with most of the appellate process. These procedures are represented by a dummy variable that indicates whether at least ten percent of the cases are decided with summary judgment procedures. In 1984, appellate systems of seven states decided from ten to fifty percent of their appeals in this manner.

---

seven range, and the research findings, of course, do not imply that large IAC's would maintain current levels of efficiency if they sat en banc.

Here again, Table 2 cannot be compared to Table 3 because the greatest changes in panel size tend to occur when IAC's are created or greatly expanded.

71. Perhaps, however, opinions take longer to write when there are larger panels because the opinion author must try to incorporate more views. Larger panels may also provide more opportunity for dissenting opinions. But these concerns probably affect only a small minority of a state's appeals, where the issues lend themselves to divergent views.

72. There were 24 within-state changes in the 15% dummy variable in Table 2, and 13 for the 50% dummy variable. Comparable figures for Table 3 are nine and five for supreme courts without IAC's, 10 and five for mid-range states, and four and three for states with large IAC's. Statistics for the percentage of cases argued are available for 30 of the 44 states. A regression containing only these states showed that the percentage of cases argued had a significant impact on productivity (coefficient = -.0025, T-Ratio = 4.00).

In the overall regression (Table 2), this variable shows no impact on output. The same result was obtained in Table 3 for supreme courts (five of which adopted summary judgment procedures). The final two analyses show barely significant impacts, in opposite directions. These results, however, are incomplete because each analysis contained only one state that adopted summary judgment procedures.

IV. CONCLUSIONS

The volume of decisions by state appellate courts has grown dramatically, as has the judges' productivity. Exploring the reasons for this growth is difficult because numerous interacting factors influence output and productivity. This research uses the only feasible research design for such a task, the multiple time-series design.

The analysis found that changes intended to increase appellate court productivity usually have that effect, although the impact is often not substantial. Importantly, the study found that adding judges ordinarily produces a corresponding increase in the number of appeals decided (provided, of course, that filings also increase). Supreme courts, however, may exhibit slight "declining returns to scale." Assigning temporary judges to appellate courts also helps increase the volume of appeals decided.

The most effective procedural changes are those that reduce the burden of writing and publishing opinions. Deciding cases without opinion greatly increases court output, and reducing the number of opinions published has a lesser, but still substantial, impact. Issuing unsigned memorandum or per curium opinions, rather than full signed opinions, has but a slight impact.

Reducing the number of cases argued also has a moderate impact on productivity. In conducting our research, we were not able to gather information concerning the actual length of oral arguments, but we determined that rule changes that reduce time limits for arguments have no impact.

The fact that curtailing opinion writing and publication, and limiting arguments increase appellate court efficiency does not, of course, automatically lead to a recommendation that these practices be adopted. Many have argued that such changes reduce the quality of justice provided by the courts. The present research looks only at the impact of changes on the number of cases decided per judge. Before judges or others can apply these

findings, they must factor in their feelings about the potential impact on the quality of appellate justice.

The act of creating intermediate courts, or expanding their jurisdiction, in itself increases the productivity of the entire appellate system. This occurs even after controlling for contemporaneous changes, such as adding judges, using three-judge panels, and routing minor appeals to the appellate courts rather than to general jurisdiction trial courts.

Changes found to have very little or no impact include reducing panel size, adopting summary procedures, and adding staff attorneys. Adding law clerks has a small impact. Appellate courts commonly make these changes in the belief that they will help them cope with the rising caseloads. The research here suggests otherwise, and judges are advised to consider other changes, especially reducing arguments and full opinions, when desiring to increase productivity. There are, however, reasons for making changes other than productivity concerns—changes that are beyond the scope of this research. For example, law clerks and staff attorneys may well increase the quality of information used by judges when making decisions.

When all is said and done, the greatest determinant of the number of decisions is the number of filings. Caseload pressures are by far the most important stimulus increasing productivity, and the great increase in decision output and productivity of appellate courts in recent decades is largely due to an equally great increase in appeals filed. The implication is that when caseloads rise, judges either work harder or reduce the amount of attention given each case in ways not measured here, such as by spending less time reading briefs and transcripts. On the other hand, filing growth does not lead to a one-to-one growth in decisions; for example, a ten percent filing increase typically stimulates about six percent more decisions. Therefore, unless more judges are added or efficiency measures adopted, the average appellate court will soon fall hopelessly behind.