

Taking and Testing Jurisprudential Regimes Seriously: A Response to Lax and Rader

Herbert M. Kritzer University of Minnesota Law School
Mark J. Richards Grand Valley State University

Lax and Rader critique our use of the Chow test in our series of articles on jurisprudential regimes on the grounds that individual justices votes are not statistically independent, which constitutes a violation of assumptions underlying the Chow test. In this response we point out that the Chow tests constituted only one part of our analysis; we also conducted a sensitivity analysis to look at the strength of the Chow tests compared to other sequential splits. Most importantly, we required that the observed statistical patterns of change be theoretically consistent with the legal changes made by the regime changing decisions; we note two areas where we did preliminary analyses that produced statistically significant results, but where those results did not make sense in light of the jurisprudence. We repeat both our Chow tests and individual interaction tests taking into account the clustering of observations. Our reanalysis provides support for some, but not all, of our original results.

Just as there has been careful examination of the underpinnings of the attitudinal model, it is fitting that our work on the relevance of law in Supreme Court decision making be subject to close reexamination. We welcome the analysis undertaken by Lax and Rader and appreciate having the opportunity to respond.¹

The authors have reanalyzed three of the four data sets in our series of articles on jurisprudential regimes with an eye to assessing whether *one* element of the analysis we conducted, the *Chow* tests, might have led us to be overconfident in the conclusions that we drew regarding whether the structure of decision making by Supreme Court justices changed in a way consistent with jurisprudential regimes.² The thrust of their argument is that because of non-independence among the votes of the individual justices in the individual cases, the statistical assumptions required for the Chow test are not met, and that a randomization methodology provides a more accurate basis of inference.

Lax and Rader focus on only one element of what we did: the Chow tests. As they note, our analysis had three separate components: an assessment of whether

the patterns of difference we observed made theoretical sense, the Chow tests, and a comparison of the statistical test for the theoretical break to other alternative breaks (sensitivity analysis). What was perhaps not clear in our presentation is the importance of the first step, which we detail below. The Chow test comparing the two time periods is by no means the definitive element of our examination of whether Supreme Court justices' decisions in an area of law could be characterized as being structured by a jurisprudential regime because many other splits would approximate our hypothesized split. It was not surprising to see statistical differences in the authors' "odd experiment." Various splits could be statistically significant, even without a basis in theory. This was why we conducted the sensitivity analysis to see whether the split we hypothesized was among the strongest *sequential* splits that we could check. Note that we did not look at all possible sequential splits; only at alternative sequential splits demarcated by years or terms of the Court. If we had looked at all splits, it is likely that a fair number would have had higher Chow test chi squares than did the theoretical split. We have no information on what proportion of

¹Jeff Lax shared a draft with us, and we were able to comment at an earlier stage.

²Lax and Rader note that they were able to replicate the results for our analyses with one exception. They note (footnote 9) that we verified that there was an error in that analysis. Kindly, they describe it as a "printing error." In fact, for the analysis in question (for the justices on the Court at the time of the *Lemon* decision), we made an error in selecting which justices to include. In the correct analysis, there was *not* a significant Chow test comparing the before/after period. The correct results are available from the senior author.

the random splits Lax and Rader examined had a substantial alignment with the theoretical split we examined, or whether those random splits that were stronger than our Chow test tended to be among those with an alignment resembling our theoretical split.

Lax and Rader justify the randomization test on the grounds that the observations are not statistically independent. As they note, in our latest article on the *Chevron* regime we did employ clustered standard errors, although those standard errors did not factor into the Chow test that we reported (Richards, Smith, and Kritzer 2006). Using Stata one can request that the variance-covariance matrix for the estimators (not just the standard errors) take into account clustering. One can then use these results to test a general linear hypothesis (Goldberger 1991, 233–37) that the coefficients as a set are equal. One issue is whether one should cluster by case or by justice, as it is not possible to do both simultaneously. We believe that the appropriate method is by case because many more predictor variables cluster within case.

We went back to our data and applied the general linear hypothesis approach to our four data sets.³ The first column of the Table 1 shows the Chow test results as reported in our original articles, both for all justices and for the justices serving on the Court at the time of the hypothesized regime change.⁴ The second column shows the results of the general linear hypothesis test without applying clustering. The third and fourth columns show the general linear hypothesis test with clustering, by justice and by case. Nothing in these alternative tests changes our conclusions.⁵

Even more important to us was whether patterns we observed in the actual coefficients made theoretical sense, based on the jurisprudential regimes. In our analysis of Establishment Clause cases, we found that the elements of the *Lemon* test were significant in the expected direction post-*Lemon* while this was not true pre-*Lemon* (Kritzer and Richards 2003). We also conducted an analysis of confessions cases.⁶ That analysis produced a Chow test that was statistically

³The test becomes what Greene (1993, 131–33) labels a Wald test.

⁴As noted previously, Lax and Rader discovered an error we had made in our published analysis for the *Lemon* Court justices. We show the correct Chow test result in Table 1.

⁵We do acknowledge that the corrected version of the analysis of the *Lemon* Court justices is different here, but that reflects the error we made in our original analysis. The lack of significance here reflects in part the very small number of pre-*Lemon* votes by this set of justices.

⁶The *Miranda* data were generously made available by Sara Benesh.

significant, but the changes in the key individual variables in the model were opposite of what we had hypothesized.⁷ Similarly, we looked at data on obscenity and commercial speech.⁸ Again, we found highly significant Chow tests, but patterns of change that did not make theoretical sense, and we did not pursue any of these analyses further.

Lax and Rader contend that when evaluated according to their randomization test, most of the significant before/after differences in individual variables disappear, and all differences associated with regimes disappear when controlling for change in personnel. As described above, Stata can take into account clustering of standard errors. As an alternative to the randomization test of Lax and Rader, we used this method to assess the significance of before/after differences in individual coefficients, by interacting the individual variables with a dummy variable representing whether the case was decided before or after the regime was established. We ran two separate analyses to control for clustering by case and by justice.

Examining our article on freedom of expression (Richards and Kritzer 2002), the key before/after difference was in the effect of the variable measuring whether government regulated speech in a content-based manner. Using their randomization test, Lax and Rader found this difference to be significant in the model with all justices, but not so in the model that controls for change in personnel. Accounting for clustering, and regardless of personnel change, we found that the p-value for the coefficient of the content-based variable was significant at the $p < .05$ level or better. We also found significant before/after differences in the effect of the content-neutral variable. (We had not hypothesized and did not find the threshold not met variable to be conditioned by the regime.)

Looking back at our research note on Establishment Clause cases, we were only able to find a significant before/after difference in one of three variables hypothesized to be of interest on the basis of the *Lemon* regime: government monitoring. Lax and Rader found a similar pattern, except that even this difference disappeared after controlling for change in personnel.

⁷Pre-*Miranda*, there was essentially a totality of circumstances test, which led us to expect variables more variables to be important before *Miranda*; in fact, none of the variables other than justices' attitude was significant. After *Miranda*, all were significant, and all but justices' attitude had changed significantly from before.

⁸The data on obscenity cases were generously made available by Kevin McGuire.

TABLE 1 Chow Tests and General Linear Hypothesis Tests

	Original ^a	Unclustered GLH ^b	Clustered by Justice GLH	Clustered by Case GLH
Freedom of Expression	124.68***	108.11***	722.77***	46.48**
Richards and Kritzer (2002)	<i>113.16***</i>	<i>85.65***</i>	<i>137.50***</i>	<i>44.14**</i>
Establishment Clause	35.29***	26.81***	24.11***	70.06***
Kritzer and Richards (2003)	<i>11.99[#]</i>	<i>6.90</i>	<i>17.41**</i>	<i>6.58</i>
Search & Seizure	72.98***	65.55***	156.62***	30.73**
Kritzer and Richards (2005)	<i>46.71***</i>	<i>43.29***</i>	<i>72.35***</i>	<i>37.68***</i>
<i>Chevron</i> Richards,	38.11***	107.02***	109.98***	33.55**
Smith, and Kritzer (2006)	<i>42.26***</i>	<i>109.21***</i>	<i>110.63***</i>	<i>48.31***</i>

Note: Analysis for all justices shown in Roman; analysis for justices on the Court at the time of the regime change shown in italics.

^aWith one exception, these are from our previously published articles. We made an error in the analysis of the *Lemon* Court justices in our Establishment Clause study; what we show above is the corrected result for that analysis.

^bGeneral Linear Hypothesis

[#]p < .10; *p < .05, **p < .01, ***p < .001

When we controlled for clustering, we confirmed their findings.

In our article on search and seizure cases (Kritzer and Richards 2005), we had hypothesized significant before/after differences for seven individual variables and had found support for six of those hypotheses, regardless of personnel change. Employing the randomization test, Lax and Rader found no evidence that differences in any of these variables were significant. (There were additional variables we had tested but had not hypothesized and did not find that the regime mattered for these variables; Lax and Rader included these variables in their Table 1.) When we accounted for clustering by case, we found support for four of the six previously supported hypotheses, but only two when we controlled for personnel change. When we accounted for clustering by justice, we found support for five of six hypotheses, regardless of personnel change.

In our article on administrative law, we had expected the justices would treat differently two jurisprudential variables, rulemaking and statute length (as a proxy for whether Congress had legislated precisely) after the *Chevron* regime was established. We reported robust standard errors that accounted for clustering in that article, but not in the form of the before/after interaction test we discuss here. For this response, when we controlled for clustering by case in the model with all justices, we did not find support for the rulemaking variable but we did find support for the statute length variable. When controlling for clustering by justice, we found significant before/after differences for both jurisprudential variables in the model that included all justices. In the models that controlled for personnel change, we found significant

differences for both variables, accounting for clustering.

Looking at critical individual variables that we hypothesized were associated with jurisprudential regimes in our articles on freedom of expression, Establishment Clause, search and seizure, and administrative law, although the randomization test accurately noted the presence and absence of some significant differences, our analyses taking into account clustering show some critical significant differences that the randomization test missed.

The broader question concerns what the randomization test may tell us about Supreme Court behavior more generally. Is there a broad pattern that specific, theoretically derived time breaks are not all that different from randomly chosen time breaks?⁹ Lax and Rader's analysis raises questions about whether that behavior is at all consistent vis-à-vis factual elements. One would expect that if one took a set of cases where there was no hypothesized change and randomly split the cases into two groups, the patterns of coefficients for the two groups would not differ in statistically significant ways. The results presented by the authors suggest otherwise. We question the conclusion of the authors that it may not make sense to use either our version or their version of the regimes test "if most Supreme Court cases are meant to be, and likely are, regime changing." The authors offer no evidence to support this assumption. More

⁹Only recently have studies of the justices' votes on the merits employed robust standard errors. Given the differences we found with and without clustering, one has to ask whether results reported in earlier studies that did not take into account clustering would hold up.

importantly, it is only through inquiry, critical analysis and debate among scholars that we can look for patterns and accept or reject them as likely or not. In this spirit, we appreciate the challenges the authors have set out here.

Finally, we offer a word about theory. In theorizing about the role of law and jurisprudential regimes in Supreme Court decision making, we have been attentive to interpretivist concerns about language (Gillman 1999). In describing how law matters, we have avoided using phrases such as whether it is likely or not that “precedents ‘bind’ future votes,” used by the authors. The justices decide cases; the law does not act independently of the justices or dictate results. Martin Shapiro’s (1964) concept of “political jurisprudence” reminds us although the justices take into account jurisprudence, they do so as political actors who are part of a political and institutional structure of government. We would certainly recoil at any conception of Supreme Court decision making that reduces the justices to robots sorting cases into bins. By focusing solely on the statistical tests, we miss some of the nuance and complexity of jurisprudential regime theory, such as using interpretive methods to trace the origins of the regimes and looking to legal scholars for confirmation of our hypotheses about the parameters of the regimes. We agree with the authors’ statement that, “Saying that votes are relatively predictable is not the same as saying we can figure out the full underlying doctrine by using statistical fact pattern analysis.” Hopefully the statistical analyses can reveal patterns but as law professors like to remind their students, there is no substitute for reading the cases.

Acknowledgments

The collection of data used in the analyses reported in this response was supported by the University of Wisconsin Political Science Department, the University of Wisconsin Graduate School, and the Univer-

sity of Wisconsin Law School. We would also like to thank Jeffrey Segal, Joseph Ignagni, Kevin McGuire, and Sara Benesh, all of whom generously made available to us data that they had collected.

Manuscript submitted 25 February 2009

Manuscript accepted for publication 7 July 2009

References

- Gillman, Howard (1999) “The Court as an Idea, Not a Building (or a Game): Interpretive Institutionalism and the Analysis of Supreme Court Decision-Making.” In *Supreme Court Decision-Making: New Institutional Approaches*, ed. C. W. Clayton and H. Gillman. Chicago: University of Chicago Press, 65–87.
- Goldberger, Arthur S. (1991) *A Course in Econometrics*. Cambridge, MA: Harvard University Press.
- Greene, William H. (1993) *Econometric Analysis*. 2nd ed. New York: Macmillan Publishing Company.
- Kritzer, Herbert M. and Mark J. Richards (2003) “Jurisprudential Regimes and Supreme Court Decision Making: The Lemon Regime and Establishment Clause Cases.” *Law & Society Review* 37 (4): 827–40.
- Kritzer, Herbert M. and Mark J. Richards (2005) “The Influence of Law in the Supreme Court’s Search-and-Seizure Jurisprudence.” *American Politics Research* 33 (1): 33–55.
- Richards, Mark J. and Herbert M. Kritzer (2002) “Jurisprudential Regimes in Supreme Court Decision Making.” *American Political Science Review* 96 (June): 305–20.
- Richards, Mark J., Joseph L. Smith, and Herbert M. Kritzer (2006) “Does *Chevron* Matter?” *Law and Policy* 28 (October): 444–69.
- Shapiro, Martin (1964) *Law and Politics in the Supreme Court*. London: Free Press of Glencoe.

Herbert Kritzer is Marvin J. Sonosky Chair of Law and Public Policy at the University of Minnesota Law School, Minneapolis, MN 55455.

Mark Richards is Associate Professor and Chair of Political Science at Grand Valley State University, Allendale, MI 49401.