

SUPPLEMENTAL WEB APPENDIX

Supplemental Appendix 1. Independence of Ratings Sources

Item Response Theory (IRT) requires that the source variables are independent of each other and some debate exists about whether the disaggregated Collins and Cooper (2012) *CSI* index values are truly independent of each other. The *CSI* index is a pseudo ordinal index that taps into four newspapers' coverage / placement of stories:

(0): Indicates no coverage anywhere in the paper

(1): Indicates that the decision was covered somewhere in the paper, but not on the front page.

(2): Indicates that the decision was covered on the front page of the paper

For this project we disaggregated the individual paper scores to create dichotomous identifiers for whether the case was discussed on the first page (i.e., a *CSI* code of 2 *versus* 1 or 0) or whether the case was discussed elsewhere but not on the first page (i.e., a *CSI* code of 1 *versus* 2 or 0). Because of the overlap with the Epstein's *NYT* measure we did not include the individual paper score for the first page of the *New York Times* in our model specification.

The coding of these different dichotomous variables is independent in terms of measurement structure. They either come from a separate independent rating source (i.e., alternative published accounts of salience or import) or they identify separate levels of import associated with the case (i.e., references to the case are either found on page 1 (or somewhere thereafter); references to the case either identify the case as a principal citation (or a regular citation)).

Theoretically it may be that the underlying salience or coverage of page 2 and thereafter articles are not strictly independent of page 1 articles. It may be some level of endogeneity exists between rater sources or between editorial page placements, thus they are not purely independent. For example, the *New York Times* coverage could be driving coverage found in competing papers or even affect the retrospective assessments of import for these cases. We cannot com-

SUPPLEMENTAL WEB APPENDIX

pletely isolate the ratings sources from each other within the context of this project, but we do follow an accepted practice within the extant literature (Clinton and Lapinski 2006). These estimates exhibit properties of scale reliability (*see*, Table 5 of manuscript) and the tests of construct validity (*see*, Table 6 of manuscript) suggest that there is value in including more ratings sources rather than fewer.

SUPPLEMENTAL WEB APPENDIX

Supplemental Appendix 2. Issues of Missing Data and the Retrospective Only Estimates

One issue that may be a source of some concern for the validity of the Judicial Accomplishment Estimates (*JAE*) is the lack of contemporaneous indicators of case salience (Esptein and Segal 2000; Collins and Cooper 2012) in the earliest era of the sample (*i.e.*, 1899 to 1946 terms). In fact, it was suggested that we limit the sample period for the estimates to the era in which contemporaneous indicators were available. We understand the concern over the missing contemporaneous ratings, but believe the answer cannot be found in the limitation of the sample period.

First, such a stance would restrict these landmark ratings to the period between 1946 and 2004 (*i.e.*, 59 term obs, which would be halved when we use the biannual congress as the unit of analysis). Given that the Clinton and Lapinski (2006) scores currently drop off at 1994 (49 term obs), that approach would make these estimates unsuitable for understanding temporal changes in inter-branch influences/constraints upon policymaking. We really require a more extended period to understand this inter-branch policy-making. As currently situated, these measures will provide coverage from 1899 to 1994. That is a period that is useful to understand majoritarian influences in the economic policy area, and in terms of civil liberties and rights issues, and it provides 95 annual observations for both the Congress and the Court.

We also do not believe that extending the contemporaneous measurement strategy (*e.g.*, the New York Times measure) into earlier eras is viable (*see*, endnote 1 of the manuscript) because of the changing nature of media coverage and the altered role of judicial branch policy-making over the sample period. Instead of discarding the 47 years of case level estimates, we considered other alternatives and decided that the best path was to evaluate the magnitude of the threat and determine whether it was problematic or not. To accomplish this task we have run an alternative

SUPPLEMENTAL WEB APPENDIX

set of estimates that only tap into the retrospective identifiers that are available throughout the early and late sample period (*see*, Table A1 and Table A2 that follow).

With these alternative estimates, we were able to look for evidence of bias related to the lack of contemporaneous estimates in the earliest period of the sample. The new table of estimates shows remarkable consistency (other than a change in scale attributable to the altered model specification). We then conducted tests of validity (*see*, Table 5 and 6 of the manuscript) that compared these original estimates versus the new retrospective estimates. The results of the validity test were likewise supportive of the original model specification with the contemporaneous rating sources. These results show that the original estimates are preferred in each of the three common sample periods (1979 to 2004; 1953 to 1999; and 1899 to 1999). If for some reason the lack of contemporaneous measures in the early sample were adverse to the validity of the case level estimates, then we would expect to find some inconsistencies across these three different sample periods.

Perhaps more important for understanding the usefulness of this new measurement strategy the alternative retrospective measure performed exceedingly well too. They were the second best performing measure in the shortest period (1979 to 2004), essentially tied for second place in the middle range sample (1953 to 1999) and it was clearly the next best performing measure in the longest sample period (1899 to 1999).

We also tested the two measures in the sample period between (1899 to 1945) when the contemporaneous ratings were not available. The model results do show that the retrospective only model minimizes the likelihood in that period, but the difference is less than 1 point on the likelihood value and the difference between the two measures in this era is not substantively meaningful.

SUPPLEMENTAL WEB APPENDIX

	est.	(s.e.) ^d	LL	LLR Test ^f
<u>1899 to 1945 terms</u>				
JAE	.48	(.03)	-6219.13	1.58
JAE – Retrospective	.45	(.02)	-6218.34	–

We find a remarkable amount of consistency and validity of these estimates both with and without the contemporaneous rater sources. That suggests the threat of missing data to the validity of the measures is reasonably minute. We make both the original and new retrospective JAE estimates available to the research community. Those scholars who remain concerned with missing data for the contemporaneous raters may adopt the retrospective only estimates instead of the original estimates.

SUPPLEMENTAL WEB APPENDIX

Table A1. MCMC estimates of ratings parameters – judicial accomplishment retrospective.

		Discrimination		Threshold	
		β^{adj}	(s.e.)	α^{adj}	(s.e.)
<u>Retrospective – Lists & Ref.</u>					
LexisNexis Landmarks		12.67	(1.91)	-.50	(.03)
CQ Guide	Savage (2010)	11.48	(1.42)	-.78	(.05)
Landmark Decisions	Finkelman & Urofsky (2008)	9.12	(1.09)	-.92	(.06)
Encyclopedia	Dawson (2001)	16.74	(2.22)	-.80	(.05)
Oxford Guide	Hall (1997)	8.18	(.99)	-.74	(.05)
Palgrave Economics & Law	Newman (1998)	2.51	(.34)	.20	(.13)
Compendium	Epstein, et al. (1994)	9.03	(1.12)	-.74	(.05)
USSC & Constitution	Kutler (1984)	10.16	(1.37)	-.62	(.04)
<u>Retrospective – Undergraduate</u>					
CLCA-Inst. Power	Epstein & Walker (2011a)	7.21	(1.03)	-.37	(.04)
CLCA-Inst. Power Appx.	Epstein & Walker (2011a)	4.70	(.72)	-.04	(.10)
CLCA-Rights	Epstein & Walker (2011b)	13.65	(2.07)	-.51	(.03)
CLCA-Rights Appx.	Epstein & Walker (2011b)	6.03	(.76)	-.50	(.03)
CLP-Power & Govt.	O'Brien (2011a)	7.65	(1.07)	-.42	(.03)
CLP-Civil Rights	O'Brien (2011b)	7.72	(.97)	-.58	(.03)
<u>Retrospective – Law School</u>					
Constitutional Law-Principal	Barnett & Katz (2013)	10.49	(1.46)	-.53	(.03)
Constitutional Law-Reg.	Barnett & Katz (2013)	7.72	(1.80)	-.07	(.14)
Constitutional Law-Principal	Sullivan & Feldman (2013a)	9.31	(1.26)	-.52	(.03)
Constitutional Law-Reg.	Sullivan & Feldman (2013a)	4.99	(.58)	-.75	(.05)
Criminal Procedure-Principal	Israel, et al. (2012)	3.32	(.47)	.09	(.11)
Criminal Procedure-Reg.	Israel, et al. (2012)	2.44	(.31)	-.09	(.07)
Criminal Procedure-Principal	Dressler & Thomas (2012)	3.65	(.50)	-.00	(.09)
Criminal Procedure-Reg.	Dressler & Thomas (2012)	2.39	(.31)	.12	(.11)
Employment Discrim.-Principal	Zimmer, et al. (2013)	2.71	(.49)	.83	(.32)
Employment Discrim.-Reg.	Zimmer, et al. (2013)	2.00	(.27)	.63	(.20)
Federal Admin. Law-Principal	Lawson (2013)	1.99	(.33)	1.32	(.38)
Federal Admin. Law-Reg.	Lawson (2013)	1.39	(.19)	1.23	(.30)
First Amdt. Law-Principal	Sullivan & Feldman (2013b)	8.72	(1.34)	-.37	(.04)
First Amdt. Law-Reg.	Sullivan & Feldman (2013b)	4.37	(.52)	-.48	(.03)
Labor Law-Principal	Cox, et al. (2011)	1.18	(.13)	2.86	(.46)
Labor Law-Reg.	Cox, et al. (2011)	.95	(.09)	3.02	(.42)

Posterior estimates are obtained through a five-chain 50K MCMC iteration with the first 40K samples burned. In total 50K posterior samples were used and the resulting posterior estimates are for adjusted parameters.

SUPPLEMENTAL WEB APPENDIX

Table A2. MCMC estimates of hierarchical parameters.

	est.	(s.e.)
General Citation Rate _(5yr)	.07	(.010)
Distinguishing Citation Rate _(5yr)	.02	(.005)
Following Citation Rate _(5yr)	.01	(.005)
Overtured Case Indicator _(0,1)	.26	(.040)
Tau adjusted _(τ adj)	.30	(.124)
Constant Value	-1.83	(.160)

Posterior estimates are obtained through a five-chain 50K MCMC iteration with the first 40K samples burned. In total 50K posterior samples were used and the resulting regressors are expressed as standardized coefficients with the exception of the dichotomous overtured case value.

SUPPLEMENTAL WEB APPENDIX

Appendix 3. Selection of Law School Texts

One issue that came up during the review process was the representativeness of the law school texts included within the analysis. We did not really approach this as an issue of representativeness of the law school texts themselves because that information would ultimately be available through an assessment of the discrimination and threshold parameters for each rating source. We were, however, concerned about the representative breadth of issue coverage to make sure we were capturing the different strains of law.

We put out feelers with 4 or 5 different law school professors, but only had luck with two with which we had previous relationships. One, in addition to his own course textbooks, got information from the school's textbook representatives about the most common texts across specialty courses. The other utilized his course adoptions and reached out to other faculty members who taught these different courses. We sincerely appreciate the help of these law professors and textbook representatives for the help in identifying sources.

The issue for inclusion in the project was not really about representativeness of the text as much as it was the usability of the data and the format of the case index. Some texts do not include a *case citation* along with the title that can be used to identify the case. We utilized any text that had the citation identifier as it made the coding of the work much more efficient and reliable.

Parameter results for the discrimination and threshold parameters show that the works are related to the latent dimension that we pull from these data, so there are no real concerns about their individual merit. It would have been nice to have others but we thought it acceptable to insure coverage of different issue areas. The following are law school texts that were considered and utilized (in bold type) for the creation of these new estimates.

SUPPLEMENTAL WEB APPENDIX

Potential Constitutional and Criminal Law Sources (selected in bold)

Allen, Ronald J., Richard B. Kuhns, and William J. Stuntz. 1995. *Constitutional Criminal Procedure: An Examination of the Fourth, Fifth, and Sixth Amendments, and Related Areas*. 3rd ed. Boston: Little Brown.

Barnett, Randy E., and Howard E. Katz. 2013. *Constitutional Law: Cases in Context*. 2nd ed. New York: Wolters Kluwer Law & Business.

Chemersinsky, Erwin. 2013. *Constitutional Law*. 4th ed. New York: Wolters Kluwer Law & Business.

Chemersinsky, Erwin, and Laurie L. Levenson. 2013. *Criminal Procedure*. 2nd ed. New York: Wolters Kluwer Law & Business.

Dressler, Joshua, and George C. Thomas. 2013. *Criminal Procedure: Principles Policies and Perspectives*. 5th ed. St. Paul, MN: West.

Israel, Jerold H., Yale Kamisar, Wayne R. Lafave, Nancy J. King, and Eve Brensike Primus. 2012. *Criminal Procedure and the Constitution: Leading Supreme Court Cases and Introductory Text*. 13th ed. St. Paul, MN: West.

Stone, Geoffrey R., Louis M. Seidman, Cass R. Sunstein, Mark V. Tushnet, and Pamela S. Karlan. 2012. *The First Amendment*. 4th ed. New York: Wolters Kluwer Law & Business.

Sullivan, Kathleen and Noah Feldman. 2013. *Constitutional Law*. 18th ed. St. Paul, MN: Foundation Press.

Sullivan, Kathleen and Noah Feldman. 2013. *First Amendment Law*. 5th ed. St. Paul, MN: Foundation Press.

Employment & Administrative Law Textbooks (selected in bold)

Breyer, Stephen G., Richard B. Stewart, Cass R. Sunstein, Adrian Vermeule, and Michael Herz. 2011. *Administrative Law and Regulatory Policy: Problems, Text and Cases*. 7th ed. New York: Wolters Kluwer Law & Business.

Cass, Ronald A., Colin S. Diver, Jack M. Beermann, and Jody Freeman. 2011. *Administrative Law: Cases and Materials*. 6th ed. Boston: CCH Inc.

Cox, Archibald, Derek C. Bok, Robert A. Gorman, and Matthew W. Finkin. 2011. *Labor Law: Cases and Materials*. 15th ed. New York: Foundation Press.

Harper, Michael C. and Samuel Estreicher. 2011. *Labor Law: Cases, Materials, and Problems*. 7th ed. New York: Kluwer Law & Business.

Goldman, Alvin L. and Roberto L. Corrada. 2011. *Labour Law in the USA*. 3rd ed. Frederick, MD: Kluwer Law International.

Lawson, Gary. 2013. *Federal Administrative Law*. 6th ed. St. Paul, MN: West.

Rothstein, Mark A., and Lance Liebman. 2011. *Employment Law: Cases and Materials*. 7th ed. New York: Foundation Press.

Zimmer, Michael J., Charles A. Sullivan, and Rebecca Hanner White. 2013. *Cases and Materials on Employment Discrimination*. 8th ed. New York: Wolters Kluwer Law & Business.

SUPPLEMENTAL WEB APPENDIX

Supplemental Appendix 4. Construct Validity Test Model Specification

When conducting validity assessment of this measure we hazard running into a couple of different challenges. First, a theoretical conflict is created when comparing it to the extant contemporaneous measures. If we were to utilize a combined contemporaneous/retrospective measure to identify significant relationships in decision-making outcomes, or other decision-making processes, we clearly open ourselves up to challenges of tautological model specifications – post decision information is being used to explain or predict the decision. Retrospective measures obviously carry inherent problems of tautological inference and endogeneity certainly exists between the presence of dissenting opinions and the media coverage that underlies many of these indicators (Maltzman and Wahlbeck 2003). We reiterate that the intended purpose of the JAE values is to create comparative leverage versus legislative branch policy-making and merely use dissent to evaluate construct validity in a localized comparative sense.

We followed the guidance of a reviewer whom suggested that we compare our measure to others in the area of consensual norms – a place where we had a self-citation (Hendershot, Huwitz, Lanier and Pacelle 2013) and where we knew the underlying data to be reliable. This test thus runs us up against the tautological inference problem (*i.e.*, split decisions are likely to be more salient / substantively important than unanimous ones), but it does have the benefit of providing coverage for the entire era. Essentially, we knew that we would have a tautology problem but we could at least draw some comparative leverage across the different existing salience / network measures to see how they performed.

To operationalize that test we obtained the range of the most recently developed continuous measures of landmark status (*i.e.*, Fowler, et al. 2007, Black, Sorenson and Johnson 2013, and Clark, Lax and Rice 2015). These values all have different key variables and it required some

SUPPLEMENTAL WEB APPENDIX

effort to import and reconcile these data with each other. After the data were clean, we ran reliability tests using Cronbach's α . Results indicate that our new measure has scale reliability with the conventional *NYT* and *CSI* measures as well as the best performing continuous alternatives.

We then ran negative binomial regressions of the number of justice dissents in three separate time periods that align with categories of the landmark measurement strategies. We utilized a common sample of case observations and tested parsimonious model specifications to let the performance of each control variable to speak for itself as much as possible. We did, however, include a control for the number of participating justices, and a couple of regime / dummy variables within the longest sample period to control for intervention points found within the Court's norm of consensus.

We adopted this strategy because it provides a level playing field to assess the various indicators of landmark status. By utilizing each landmark indicator within the same sample of observations, we are able to understand relative differences in the ability to isolate systematic variance. This does not mean that all of these variables would be viable explanations of justice dissent. Retrospective measures obviously carry the inherent problems of tautological inference and endogeneity certainly exists between the presence of dissenting opinions and the media coverage that underlies many of these indicators (Maltzman and Wahlbeck 2003).

Again, the essential problem here is that we are not exactly clear of the tautological inference problem. With that in mind, we decided to run a very parsimonious model specification and not replicate the model specification found in some other piece of literature. First, because we didn't really have a good one to use that covered the entire sample period. Second, because we were not on firm theoretical grounds in using this post-decision rater based measure to understand behav-

SUPPLEMENTAL WEB APPENDIX

ior like coalition formation. The results simply will tell us whether we are adequately capturing variance versus other possible measurement strategies.

Following a request from reviewers, we eventually revised the validity test and moved away from the parsimonious test. It was suggested that we line up as best as possible to Hendershot, Hurwitz, Lanier and Pacelle (2013) dissent article. This was somewhat problematic too, as it had a different unit of analysis – an annual time series dependent variable and Box-Jenkins ARIMA estimation. Here we have a case level unit of analysis so the specification is slightly different.

We continued to include the control for the number of participating justices on the decision, which is necessary for the event count model to work properly. We then included a range of issue controls for: 1) criminal issues; 2) civil liberties and rights issues; 3) institutional power and federalism issues; and 4) original and miscellaneous issues. This series of controls left economic issues as the uncontrolled null specification. We then controlled for the number of cases on the Court's annual docket, the standard deviation of associate justice preferences found on the Court that term, and the mean experience level of the associate justices. Finally, we controlled for the different chief justices with dummy variables. We left the first chief justice of each sample period as the uncontrolled null specification.

The results of the revised model specifications did not show any real deviation from the earlier parsimonious model specification. After controlling for a range of alternative explanations (*e.g.*, issue types, ideological composition, experience levels on the bench, and chief justice leadership) we continue to find that our new estimates minimize the likelihood value in common sample periods from: 1) 1979 to 2004; 2) 1953 to 1999; and 3) 1899 to 1999. In addition, the alternative *JAE retrospective* estimates that we created are typically the next best performing measure.

SUPPLEMENTAL WEB APPENDIX

We completely agree with those who would suggest that these new JAE estimates are not appropriate for understanding individual justice behavior given the tautological inference issue. To some extent, the translation / replication of the dissent model specification contributes to the perception that it is useful in predicting justice behavior. We would probably prefer to use the parsimonious model specifications because we think they create less confusion with respect this problem, but have published the more intensive model specifications associated with the Hendershot, Hurwitz, Lanier and Pacelle (2013) dissent article.

SUPPLEMENTAL WEB APPENDIX

References

- Black, Ryan C., Maron W. Sorenson, and Timothy R. Johnson. 2013. "Toward an Actor-Based Measure of Supreme Court Case Salience: Information Seeking and Engagement During Oral Arguments." *Political Research Quarterly* 66(4):804-18.
- Clark, Tom S., Jeffrey R. Lax, and Douglas Rice. 2015. "Measuring the Political Salience of Supreme Court Cases." *Journal of Law and Courts* 3(1):37-65.
- Clinton, Joshua D., and John S. Lapinski. 2006. "Measuring Legislative Accomplishment, 1877-1994." *American Journal of Political Science* 50(1):232-49.
- Collins, Todd A., and Christopher A. Cooper. 2012. "Case Salience and Media Coverage of Supreme Court Decisions: Toward a New Measure." *Political Research Quarterly* 65(2):396-407.
- Epstein, Lee, and Jeffrey A. Segal. 2000. "Measuring Issue Salience." *American Journal of Political Science* 44(1):66-83.
- Fowler, James H., Timothy R. Johnson, James F. Spriggs II, Sangick Jeon, and Paul J. Wahlbeck. 2007. "Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court." *Political Analysis* 15(3):324-46.
- Hendershot, Marcus E., Mark S. Hurwitz, Drew Noble Lanier, and Richard L. Pacelle, Jr. 2013. "Dissensual Decision Making: Revisiting the Demise of Consensual Norms with the U.S. Supreme Court." *Political Research Quarterly* 66(2):467-81.
- Maltzman, Forrest, and Paul J. Wahlbeck. 2003. "Salience or Politics: New York Times Coverage of the Supreme Court." Presented at the Annual Meeting of the Midwest Political Science Association, Chicago.