B.J.Pol.S., Page 1 of 8
doi:10.1017/S000712341600079X

Notes and Comments

Are Things Really Getting Better? How To Validate Latent Variable Models of Human Rights

CHRISTOPHER J. FARISS*

Keywords: human rights; treaty compliance; measurement; latent variable model; validity

Has respect for human rights improved? The validity of inferences made about human rights, treaty compliance, and any other difficult or impossible to observe concepts depends on specifying a theoretically informed model that best approximates our understanding of the specific concept under study. New latent variable models of human rights and treaty compliance (1) gather together diverse sources of information about human right abuse, (2) assess the relative quality of that information as it relates to the underlying theoretical concept and (3) quantify the uncertainty of estimates of human rights abuse that the models generate.[1] Evidence from these models suggests that respect for human rights is improving over time and casts doubt on earlier claims that human rights treaties are associated with lower levels of respect for human rights.[2] In a recent critique of Fariss,[3] Cingranelli and Filippov[4] raise several issues regarding the estimation of the latent variable model for human rights and the use of this variable in regression models of human rights treaty compliance. Their primary critiques rest on their argument that the latent variable estimates are not valid before the Cingranelli and Richards (CIRI) human rights data series begins in 1981 because the latent variable estimates are extrapolations 'based on very sporadic and eclectic bits of information'.[5] In addition to downplaying the validity of these other human rights variables, Cingranelli and Filippov[6] present anomalous cases (for example, the United States in 1953) and replications of regression models presented in Fariss,[7] which exclude data from 1965–80. These authors also argue that only model specifications that include a measure of democracy are valid, and that changes in the number of

* Department of Political Science, University of Michigan (email: cjfariss@umich.edu; cjf0006@gmail.com).
I would like to thank Margaret Ariotti, Megan Becker, Daniel Berliner, Nicole Bonoff, Chad Clay, Charles Crabtree, Geoff Dancy, Cassy Dorff, Jesse Driscoll, Maya Duru, Kristine Eck, Daniel Enemark, James Fowler, Micah Gell-Redman, Holly Guthrey, Peter Haschke, Cullen Hendrix, Danny Hill, Alex Hughes, Michael Kenwick, Milli Lake, Yon Lupu, Jonathan Markowitz, Kyle Marquardt, Jamie Mayerfeld, Amanda Murdie, Michael Nelson, Dan Pemstein, Keith Schnakenberg, Paul Schuler, Jaime Settle, Brigitte Seim, Tara Slough, Reed Wood, and Thorin Wright for many helpful comments and suggestions. The code and data files necessary to implement the models in R are publicly available at the Harvard Dataverse Network https://dataverse.harvard.edu/dataverse/CJFariss, which can be linked to through http://cfariss.com/. The direct link to this dataverse is doi:10.7910/DVN/UIUR8J. This research was supported in part by the McCourtney Institute for Democracy Innovation Grant, and the College of Liberal Arts, both at Pennsylvania State University. Data replication sets are available at http://dataverse.harvard.edu/dataverse/BJPolS and online appendices are available at http://dx.doi.org/doi:10.1017/S000712341600079X.

[1] Fariss 2014; Fariss 2017; Schnakenberg and Fariss 2014.
[2] Hafner-Burton 2013; Hafner-Burton and Tsutsui 2005, 2007; Hathaway 2002; Hollyer and Rosendorff 2011; Posner 2014.
[3] Fariss 2017.
[4] Cingranelli and Filippov 2017.
[5] Cingranelli and Richards 1999.
[6] Cingranelli and Filippov 2017.
[7] Fariss 2017.

democratic states in the international system account for the observed patterns in the new latent human rights estimates that incorporate the changing standard of accountability. I address the critique in three parts.

First, latent variables allow for the exploration of deviant or unexpected cases (for example, the CIRI human rights data categorizes Sweden in 2011 and Guatemala in 1983 as engaging in the same level of torture). This type of case study is a productive research design strategy for identifying new theoretical concepts that relate to other sources of bias in the human rights documentary sources. To enhance validity, these theoretical concepts, like the changing standard of accountability, should be incorporated into future versions of the latent human rights model.

Second, contrary to the interpretation presented in Cingranelli and Filippov,[8] an analysis of the existing latent human rights variables and a measure of democracy reveals important new evidence in support of the relationship between the changing standard of accountability and human rights documentation. Only the human rights estimates from the changing standard of accountability model show a positive trend for democratic country-years.

Finally, new replication studies that use existing and new human rights data corroborate the findings of a positive correlation between human rights compliance and treaty ratification. The replications demonstrate that a reduction in sample size by restricting the start year to 1981 as opposed to 1980 or any earlier or later year is an arbitrary choice. The gradual change in the level of statistical significance obtained from these models is not surprising, because (1) the number of country-year units in the regression models decreases as the start year for the sample of each model increases over time and (2) a greater number of countries enter the sample having already ratified an increasing number of available human rights treaties.

EXPLORATION OF DEVIANT CASES IMPROVES LATENT VARIABLE MODELS

There are several event-based variables in the latent human rights model that are indicative of information about a specific type of repressive event. These variables, along with the standards-based human rights variables, help to inform the estimation of the country-year latent variable estimates from 1949 and now updated through 2013. For two of the five event-based variables, the United States was coded as repressive for specific reasons: the United States engaged in political killings during the 1950s and 1960s in the American South and it executed two Soviet spies in 1953. These are not trivial matters. These events do not even pick up the investigations into communist activists by Senator Joseph McCarthy that were also taking place in the early 1950s. Of course, monitors and the media may be more aware of these events because of the high levels of press freedom in the United States relative to other countries. This is an example of the challenges of modeling human rights respect that the latent variable model helps to address.

Cingranelli and Filippov[9] claim that the latent human rights variable estimates are not valid before the CIRI human rights data series begins in 1981. As evidence to support this argument and the choice to reduce the sample size of their replications of the models presented in Fariss,[10] Cingranelli and Filippov[11] select specific country-year examples (for example, the United States in 1953) that have unexpected values on the latent variable. Unfortunately, however, there is no model-free way to estimate unobservable concepts such as human rights. Even the CIRI human rights data – models that assume equal weighting of human rights indicators and no error – generate cases with unexpected values (for example, Sweden is coded as a country that tortures across many years). Latent variable models, with their focus on the theoretical relationship between data and model parameters, offer a principled way to bring together information from different documentary sources and make sense of both the individual pieces of information and the underlying theoretical concept. It is possible to use the information from these models to identify and evaluate cases with unexpected values for the latent variable (deviant cases) and to incorporate new theoretical concepts into new and updated versions of the latent variable model. A deviant

[8] Cingranelli and Filippov 2017.
[9] Cingranelli and Filippov 2017.
[10] Fariss 2017.
[11] Cingranelli and Filippov 2017.

case is an observation that is coded at a surprising value or outlier along some theoretical concept.[12] The identification of such cases does not undercut the progress already made in enhancing the validity of recent versions of the latent human rights variable, because each new model has been able to distinguish between theoretically distinct cases that earlier variables were not able to identify.[13]

The United States earns its lowest value on the latent variable because it is coded 1 for two of the events-based variables described in Fariss.[14] The United States is coded 1 by the WHPSI (*World Handbook of Political and Social Indicators*) *Political Executions* variable because of the execution of Julius and Ethel Rosenberg.[15] It was also coded 1 by Rummell's *Genocide and Politicide* variable because of politically motivated killings that occurred within the United States.[16] Why might the values for these variables produce the anomalous estimate for the United States in 1953?

For the Rummel data, a much broader definition of government killing is used compared to other similar measures included in the latent variable model,[17] such that a country is coded 1 if there is evidence that its government deliberately engaged in killing inside and outside its borders. Unlike the other measures of mass killing and genocide used in the latent variable model, the cases represented in the Rummel data might include widespread killing or killings targeted at political opponents or groups not specified in the definition of geno-politicide used by Harff[18] and massive repressive events used by Harff and Gurr.[19] This focus on gathering information about all government-sanctioned killing events could mean that there is more bias inherent in the Rummel data compared to the other event-based variables if he excluded small-scale killings in systematically more repressive or less accessible political contexts.[20]

All of the producers of the events-based variables are aware of this possibility. This is why information from multiple documentary sources is used to help code and corroborate the coding of the event counts of each case. When new information about repressive actions becomes available from NGOs, news reports, historians or truth commissions, these scholars update their data (the codebooks for each of these variables discuss these issues at length). Also for the Rummel data, there may be additional bias for certain country-years with respect to the latent estimate of human rights if the state in question only sanctioned killing during involvement in external conflicts. Rummel considers the individuals killed in the Korean War and the war in Vietnam in the counts he generates for the United States. However, he also considers deaths that occurred within the United States. To mitigate the potential threat to the validity of the latent variable estimates, binary event-based indicators were included in the latent variable model in place of the raw event counts.

For the WHPSI *Political Executions* variable, there is the potential for bias with respect to the attention and access of media source material in each country-year. For this reason, Taylor and Jodice used both international and regional sources for every case.[21] Nonetheless, they state in their codebook that systematic under-reporting from particularly repressive countries could lead to biases in the raw event counts.[22] To mitigate the threat to validity from systematically different counts, the raw count data from these reported events is collapsed into a binary variable before it is included in the latent variable model.

It is clear from this discussion that there are theoretical concepts – the changing standard of accountability, the level of media coverage or ability to gain access to a country – which might confound the relationship between the estimate of the latent human rights variable and the observed data. Each new version of the latent variable model has the potential to address one or more of the issues revealed by exploring deviant cases such as the United States in 1953. There are some biases that the latent variable

[12] Lijphart 1971; Seawright and Gerring 2008.
[13] Schnakenberg and Fariss 2014.
[14] Fariss 2014.
[15] Taylor and Jodice 1983.
[16] Rummel 1995; Wayman and Tago 2010.
[17] Harff 2003; Harff and Gurr 1988.
[18] Harff 2003.
[19] Harff and Gurr 1988.
[20] Wayman and Tago (2010) consider the number of killings threshold for inclusion in the dataset.
[21] Taylor and Jodice 1983.
[22] Taylor and Jodice 1983.

model already begins to address such as the changing standard of accountability. For example, the CIRI data categorizes Sweden in 2011 and Guatemala in 1983 as engaging in the same level of torture. The latent variable model that accounts for the changing standard of accountability makes this temporal comparison more plausible by correcting for differences in the standard of accountability over time. The investigation of other sources of potential bias in the monitoring of human rights is an active area of applied research.[23]

VALIDATING PATTERNS OF HUMAN RIGHTS AND DEMOCRACY

Another type of validity assessment involves comparing one estimated variable to another variable to which it should theoretically relate.[24] Cingranelli and Filippov[25] graph the latent human rights variables and the Polity2 democracy variable. The trends of the two latent human rights variables (constant standard and changing standard versions) for both democratic and non-democratic country-years are quite different, which Cingranelli and Filippov do not consider.[26]

That the number and proportion of democratic states is increasing in the international system over time is not contested (see the figures in the Appendix). In contrast to this general increase over time, the level of human rights remains flat or decreases for both democracies and non-democracies according to the CIRI physical integrity index, the Political Terror Scale and the latent human rights variable that assumes a constant standard of accountability (see the figures in the Appendix). For the latent variable model that assumes a constant standard of accountability, human rights have been decreasing since the early 1980s for democratic country-years and the 1950s for non-democratic country-years. However, the latent variable that accounts for the changing standard of accountability shows an increasing trend in the level of respect for human rights after a low point in the early 1990s for the democratic country-years and the mid-1970s for non-democratic country-years.

Across all human rights variables, there are clear differences in the level of respect for human rights between democracies and non-democracies. However, without the assumption of the changing standard of accountability, one must believe that the level of human rights has steadily decreased since a high point in the early 1980s. Are democracies really becoming worse and worse abusers of human rights? Probably not. What is much more likely is that the standard of accountability is improving as monitoring agencies look harder for abuse, look in more places for abuse, and classify more acts as abuse.

The latent variable model estimates that incorporate the concept of the changing standard of accountability are more related to democracy than the latent variable estimates that do not incorporate this concept.[27] However, the strength of these relationships raises an important theoretical issue that complicates regression-based analysis with these variables, which Cingranelli and Filippov[28] do not consider. Hill[29] and Hill and Jones[30] analyzed the conceptual and operational overlap between measures of human rights and democracy. These authors demonstrate that researchers need to exercise caution when evaluating the empirical relationship between these two variables. Though a measure of democracy is included in the regression models presented in Fariss,[31] this is not the case for every model specification because of the potential bias caused by the operational overlap between the *Human Rights* dependent

[23] Bagozzi and Berliner 2015; Clark and Sikkink 2013; Dancy and Fariss 2017; Eck and Fariss 2017; Fariss et al. 2015; Fariss and Dancy 2017; Hill, Moore, and Mukherjee 2013.

[24] Adcock and Collier 2001; Trochim and Donnelly 2008.

[25] Cingranelli and Filippov 2017.

[26] Cingranelli and Filippov 2017.

[27] Correlation coefficients between the Polity2 variable and the two competing latent variables quantify these visual patterns: 0.374 [95% *Credible Interval*: 0.367, 0.380] for the constant standard model and 0.454 [95% *Credible Interval*: 0.448, 0.460] for the changing standard model. The correlation coefficient between Polity2 and the CIRI physical integrity index is 0.391 [95% Confidence Interval: 0.365, 0.416].

[28] Cingranelli and Filippov 2016.

[29] Hill 2016.

[30] Hill and Jones 2014.

[31] Fariss 2017.

variable and the *Democracy* independent variable. As discussed in Fariss,[32] though the individual model coefficients vary, the differences between these coefficients are consistent across model specifications.

THE ASSOCIATION BETWEEN TREATY RATIFICATION AND COMPLIANCE

Human Rights and Treaty Ratification Replication Analysis

Cingranelli and Filippov[33] replicated the regression models presented in Fariss[34] after first removing data from 1965–80. Generally, when the number of units in a statistical test is reduced, the standard errors for the estimates increase because the standard errors are a function of the sample size *n*. The truncation of the sample to 1981 excludes country-years that had already had the opportunity to ratify available human rights treaties, which complicates the analysis of pooled cross-sectional time-series data.[35] To justify their truncation decision, Cingranelli and Filippov[36] suggest the latent variable estimates are not valid before the CIRI human rights data series begins in 1981. As already discussed, this is an unfounded criticism of the latent variable model and the other human rights variables that enter the model prior to 1981, which includes the Political Terror Scale available starting in 1976,[37] a measure of genocide starting in 1956,[38] a measure of massive repressive events beginning in 1945,[39] a measure of democide/politicide beginning in 1949[40] and a measure of political executions beginning in 1948.[41] These are not just 'eclectic bits' of data but well documented and reliable indicators of repression, which are available for many of the country-year units that enter the model (see the Appendix for the temporal availability of these data). Reducing the sample size in 1981 as opposed to 1980 or earlier is an arbitrary decision, which becomes obvious when considering the average level of uncertainty for the country-year latent variable estimates each year.

The level of information that each observed variable brings to the estimates of the latent variable is based on the relative information content of one variable compared to all the others. A useful feature of the model, then, is that missing data does not lead to a loss of country-year observations, but only increases the uncertainty of the estimate of a given country-year, conditional on the number of indicators available for that unit and the relative information content of all the other available indicators. Boxplots in the Appendix display the distribution of latent variable standard errors for each country-year unit each year. The estimates of uncertainty – the standard deviations of the latent variable estimates – are in part a function of the number of human rights variables available for a given country-year unit, which is incorporated into the regression models presented in Fariss.[42]

In the Appendix, figures for every model specification and every sample of country-year units with a different start year (1949–2010) show the coefficient estimates for the two competing regression models (upper and middle panels) and the differences between the coefficients from these models (lower panel). The figures represent sixty-two samples (the start year for each sample increases from 1949 through 2010), for two competing dependent variables, for eight different regression model specifications, for ten different treaty variables, or $62 \times 2 \times 8 \times 10 = 9{,}920$ regression models. When estimating these models, the standard errors increase slightly as units from earlier years are removed from the sample each year. As the start year for these samples enters the early to mid-1970s to mid-1980s, the difference between the coefficients

[32] Fariss 2017.
[33] Cingranelli and Filippov 2017.
[34] Fariss 2017.
[35] The first UN human rights treaty was open for signature in 1965 and came into force in 1969 (International Convention on the Elimination of All Forms of Racial Discrimination).
[36] Cingranelli and Filippov 2017.
[37] Gibney, Cornett, and Wood 2012.
[38] Harff 2003.
[39] Harff and Gurr 1988.
[40] Rummel 1995; Wayman and Tago 2010.
[41] Taylor and Jodice 1983.
[42] Fariss 2017.

begins to become statistically indistinguishable from 0. However, the regression coefficients from the two competing models also become statistically indistinguishable from 0. The eventual lack of statistical significance is not surprising because the number of units is decreasing and, as the start year for the sample increases, more countries enter the sample having already ratified an increasing number of treaties. Conditional on the number of country-year units in the model, there is either (1) a significant, positive relationship between treaty ratification and human rights compliance or (2) not enough data to prove either a positive or negative relationship. The results reported in Fariss et al. and Hill and Jones[43] directly contradict the negative correlations reported in earlier studies[44] and cast considerable doubt on studies that begin with this negative correlation as a puzzle that needs to be explained.[45]

New Human Rights Data and Treaty Ratification Replication Analysis

The best strategy for assessing the validity of an inference is to replicate the test using new data. New, expert-coded human rights indicators were recently published as part of the Varieties of Democracy (V-DEM) Project.[46] This project uses multiple coders per country-year unit (at least five coders per unit, but often many more) to generate latent scores based on categorical questions answered by each coder for each country-year item. The model accounts for disagreement between coders, and generates measurements of uncertainty conditional on the number of (and agreement between) coders as well as coder reliability over time. The V-DEM team has coded several human rights variables, two of which are physical integrity variables: (1) freedom from political killing and (2) freedom from torture (see the Appendix).

Unlike the standards-based human rights data, the V-DEM project controls the standards used to assess each of their variables (that is, the questions wording and format, which is displayed in the Appendix). Moreover, because the coders have completed the questions over the relatively short time span of the past four years, it is unlikely that the V-DEM human rights scores are temporally biased in the same way as the standards-based human rights data. That is, unlike the human rights reports, the V-DEM data are based on question responses that are *produced* consistently with respect to time. Like the event-based data, however, the V-DEM expert coders rely on their knowledge of evidence from the historical record. As the deviant case of the United States in 1953 illustrates, the historical record provides different levels of information for certain cases. These differences may lead to biased responses from some of the coders if they do not have access to relevant information about the specific country-year case. Though the V-DEM measurement model attempts to address the disagreement between coders, bias might still persist if the expert coders are using the same historical source material. The exploration of these potential biases and how they relate to the biases from the standards-based and events-based data are important areas of research that will inform new versions of the latent human rights model.

Figures in the Appendix plot the yearly average for the two V-DEM human rights variables from 1949–2013. These visualizations show very similar upward trends in respect for human rights after the end of Cold War, which is consistent with the pattern of the latent variable that accounts for the changing standard of accountability. Also in the Appendix, replications figures present coefficients for sixty-two samples (the start year for each sample increases from 1949 through 2010), two new V-DEM human rights dependent variables, eight different regression model specifications and four different treaty variables, or $62 \times 2 \times 8 \times 4 = 3{,}968$ regression models. The results from the V-DEM replication models corroborate the positive correlation found between human rights compliance and treaty ratification reported in Fariss and replicated above.[47]

[43] Cingranelli and Filippov 2017; Fariss 2014, 2017; Hill and Jones 2014. Hill and Jones (2014) validate the finding that ratification of the Convention Against Torture and the International Covenant on Civil and Political Rights are positively associated with respect for human rights.

[44] Hafner-Burton and Tsutsui 2005; Hafner-Burton and Tsutsui 2007; Hathaway 2002.

[45] Hafner-Burton 2013; Hollyer and Rosendorff 2011; Posner 2014.

[46] Coppedge et al. 2014; McMann et al. 2016; Pemstein, Tzelgov, and ting Wang 2015; Pemstein et al. 2015.

[47] Fariss 2017.

CONCLUSION

While Cingranelli and Filippov[48] argue that respect for human rights is declining and unaffected by treaty ratification, this claim is not supported by the available empirical evidence. Since the end of World War II, state officials have been signing and ratifying an increasing number of UN human rights treaties. Over the same period of time, monitoring organizations have been looking harder for abuse because of more and better information, looking in more places for abuse with the aid of an increasingly dense network of international and domestic civil society organizations, and classifying more acts as abuse because of an increasing sensitivity to (and awareness of) the various kinds of ill treatment and abuse that had not previously warranted attention. As Sikkink notes, these organizations 'have expanded their focus over time from a narrow concentration on direct government responsibility for the death, disappearance, and imprisonment of political opponents to a wider range of rights, including the right of people to be free from police brutality and the excessive use of lethal force'.[49] These are the reasons why the standard of accountability used to produce human rights documents is becoming increasingly stringent over time, and why previous studies have discovered negative patterns instead of positive ones.

REFERENCES

Adcock, Robert, and David Collier. 2001. Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review* 95 (3):529–46.

Bagozzi, Benjamin, and Daniel Berliner. 2015. The Politics of Scrutiny in Human Rights Monitoring: Evidence from Structural Topic Models of U.S. State Department Human Rights Reports. *Political Science Research and Methods*. doi: https://doi.org/10.1017/psrm.2016.44.

Cingranelli, David L., and David L. Richards. 1999. Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights. *International Studies Quarterly* 43 (2):407–17.

Clark, Ann Marie, and Kathryn Sikkink. 2013. Information Effects and Human Rights Data: Is the Good News About Increased Human Rights Information Bad News for Human Rights Measures? *Human Rights Quarterly* 35 (3):539–68.

Coppedge, Michael, John Gerring, Stafan I. Lindberg, Jan Teorell, Daniel Pemstein, Eitan Tzelgov, Yi ting Wang, Adam Glynn, David Altman, Michael Bernhard, M. Steven Fish, Allen Hicken, Kelly McMann, Pamela Paxton, Megan Reif, Svend-Erik Skaaning, and Jeffrey Staton. 2014. V-Dem: A New Way to Measure Democracy. *Journal of Democracy* 25 (3):159–69.

Dancy, Geoff, and Christopher J. Fariss. 2017. Rescuing Human Rights Law from International Legalism and its Critics. *Human Rights Quarterly* 39 (1):1–36.

Eck, Kristine, and Christopher J. Fariss. 2017. Ill Treatment and Torture in Sweden: A Critique of Cross-Case Comparisons. Working Paper.

Fariss, Christopher J. 2014. Respect for Human Rights Has Improved Over Time: Modeling the Changing Standard of Accountability in Human Rights Documents. *American Political Science Review* 108 (2):297–318.

——. 2017. Human Rights Treaty Compliance and the Changing Standard of Accountability. *British Journal of Political Science* (Forthcoming).

Fariss, Christopher J., Fridolin J. Linder, Zachary M. Jones, Charles D. Crabtree, Megan A. Biek, Ana-Sophia M. Ross, Taranamol Kaur, and Michael Tsai. 2015. Human Rights Texts: Converting Human Rights Primary Source Documents into Data. *PLOS ONE* 10 (9):e0138935.

Fariss, Christopher J, and Geoff Dancy. Measuring the Impact of Human Rights: Conceptual and Methodological Debates. *Annual Review of Law and Social Science* 13:TBD (November 2017). Available at http://www.annualreviews.org/doi/abs/10.1146/annurev-lawsocsci-110316-113333.

Gibney, Mark, Linda Cornett, and Reed M. Wood. 2012. Political Terror Scale. Available from http://www.politicalterrorscale.org/.

[48] Cingranelli and Filippov 2017.
[49] Sikkink 2011, 159.

Hafner-Burton, Emilie M. 2013. *Making Human Rights a Reality*. Princeton, NJ: Princeton University Press.

Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. Human Rights in a Globalizing World: The Paradox of Empty Promises. *American Journal of Sociology* 110 (5):1373–411.

——. 2007. Justice Lost! The Failure of International Human Rights Law to Matter Where Needed Most. *Journal of Peace Research* 44 (4):407–25.

Harff, Barabara. 2003. No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder Since 1955. *American Political Science Review* 97 (1):57–73.

Harff, Barbara, and Ted R. Gurr. 1988. Toward Empirical Theory of Genocides and Politicides: Identification and Measurement of Cases Since 1945. *International Studies Quarterly* 32 (3):359–71.

Hathaway, Oona A. 2002. Do Human Rights Treaties Make a Difference? *Yale Law Journal* 111 (8): 1935–2042.

Hill, Daniel W. Jr. 2016. Democracy and the Concept of Personal Integrity Rights. *Journal of Politics* 78 (3):822–35.

Hill, Daniel W. Jr., Will H. Moore, and Bumba Mukherjee. 2013. Information Politics v Organizational Incentives: When Are Amnesty International's 'Naming and Shaming' Reports Biased? *International Studies Quarterly* 57 (2):219–32.

Hill, Daniel W. Jr., and Zachary M. Jones. 2014. An Empirical Evaluation of Explanations for State Repression. *American Political Science Review* 108 (3):661–87.

Hollyer, James R., and B. Peter Rosendorff. 2011. Why Do Authoritarian Regimes Sign the Convention Against Torture? Signaling, Domestic Politics and Non-Compliance. *Quarterly Journal of Political Science* 6:275–327.

Lijphart, Arend. 1971. Comparative Politics and the Comparative Method. *American Political Science Review* 65 (3):682–93.

McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell, and Staffan I. Lindberg. 2016. Strategies of Validation: Assessing the Varieties of Democracy Corruption Data. Working Paper No. 23. Gothenburg, Sweden: Varieties of Democracy Institute.

Pemstein, Daniel, Eitan Tzelgov, and Yi ting Wang. 2015. Evaluating and Improving Item Response Theory Models for Cross-National Expert Sur. Working Paper No. 1. Gothenburg, Sweden: Varieties of Democracy Institute.

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi ting Wang, and Farhad Miri. 2015. The Varieties of Democracy Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data. Working Paper No. 21. Gothenburg, Sweden: Varieties of Democracy Institute.

Posner, Eric A. 2014. *The Twilight of Human Rights Law*. Oxford: Oxford University Press.

Rummel, Rudolph J. 1995. Democracy, Power, Genocide, and Mass Murder. *Journal of Conflict Resolution* 39 (1):3–26.

Schnakenberg, Keith E., and Christopher J. Fariss. 2014. Dynamic Patterns of Human Rights Practices. *Political Science Research and Methods* 2 (1):1–31.

Seawright, Jason, and John Gerring. 2008. Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly* 61 (2):294–308.

Sikkink, Kathryn. 2011. *The Justice Cascade: How Human Rights Prosecutions are Changing World Politics*. New York: The Norton Series in World Politics.

Taylor, Charles Lewis, and David A. Jodice. 1983. *World Handbook of Political and Social Indicators Third Edition*, Vol. 2, Political Protest and Government Change. New Haven, CT: Yale University Press.

Trochim, William M. K., and James P. Donnelly. 2008. *Research Methods Knowledge Base*, 3rd Edition. Mason, OH: Atomic Dog.

Wayman, Frank W., and Atsushi Tago. 2010. Explaining the Onset of Mass Killing, 1949–87. *Journal of Peace Research* 47 (1):3–13.