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

Christopher J. Fariss*†

Abstract

Has respect for human rights improved? Answering this question requires valid comparisons of repression levels in many different states across time. Theoretically informed latent variable models of difficult-to-measure concepts such as human rights abuse by state agents, prevalence of dissent among citizens, and treaty compliance tell a different empirical story of global patterns of repression and reform than models that do not employ the same methodology. The use of the methodology is controversial for this reason. This paper addresses several critiques of the use of these models. Latent variable model parameters are useful for exploring deviant or unexpected cases, which help to identify new theoretical concepts that are related to other sources of bias in the human rights documentation. Several replication tests, using existing and new global human rights data, suggest a positive correlation between human rights compliance and treaty ratification. Respect for human rights across states is improving over time.

*Assistant Professor, Department of Political Science, University of Michigan, 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, 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/. 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.

1 Introduction

Has respect for human rights improved? The validity of inferences made about human rights by state agents, prevalence of dissent among citizens, treaty compliance, and any other difficult or impossible to observe concept depends on specifying a theoretically informed model that best approximates our understanding of the specific concept under study. The importance of careful model specification is not a new one in these fields of inquiry.[1] In several recently published articles, I develop new latent human rights models to (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 of interest, and (3) quantify the uncertainty of estimates of human rights abuse that are generated from the models.[2] The evidence from these models suggests that respect for human rights across states is improving over time and casts doubt on earlier claims that human rights treaties are associated with lower levels of respect for human rights.[3] However, controversy persists over the appropriateness of these models.

In a recent paper, Cingranelli and Filippov (2016), 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. The primary critiques of Cingranelli and Filippov (2016) rest on their argument that the latent variable estimates are not valid before the Cingranelli and Richards human rights data series begins in 1981 because the latent variable estimates are extrapolations "based on very sporadic and eclectic bits of information" (pg.TBD). In addition to downplaying the validity of these other human rights variables, Cingranelli and Filippov (2016) highlight anomalous cases and insignificant regression results from certain model specifications to support their claim. Furthermore, these authors argue that changes in the number of 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. However, these authors misinterpret the positive changes in respect for human rights by both democratic and

[1] For example, Brysk (1994) makes this point in an essay in which she argues for "the importance of careful specification of the political processes being modeled through measurement and explicit justification of the use of particular measures to represent those processes" (692). Interested readers should consult the edited volume by Jabine and Claude (1992) and a symposium on the "Statistical Issues in the Field of Human Rights" published in *Human Rights Quarterly* (Vol. 8, No. 4, 1986) for early debates about these measurement issues.

[2] Fariss (2014, 2015); Schnakenberg and Fariss (2014).

[3] See for example Hathaway (2002); Hafner-Burton and Tsutsui (2005, 2007); Hafner-Burton (2013); Hollyer and Rosendorff (2011); Posner (2014); Vreeland (2008).

non-democratic states that have occurred over time when using the latent variable that accounts for the changing standard of accountability. Overall, the claims made by Cingranelli and Filippov (2016) are not supported by the empirical record.

I address the critique in three parts. First, the latent variable model parameters allow for the exploration of deviant or unexpected cases. This type of case study is a productive research design strategy for identifying new theoretical concepts that are related to other sources of bias in the human rights monitoring reports and other related sources of information. These theoretical concepts, just like the changing standard of accountability, can be incorporated into new versions of the latent human rights variable model. Second, analysis of the existing latent human rights variables and a measure of democracy reveals additional evidence in support of the relationship between the changing standard of accountability and human rights documentation. The evidence reported in Cingranelli and Filippov (2016) and in this paper suggests that human rights are improving on average in both democratic and non-democratic states. Finally, new replication studies that use existing[4] and new human rights data[5] corroborate results for a positive correlation between human rights compliance and treaty ratification.[6] Respect for human rights across states is improving over time.

2 Exploration of Deviant Cases Improves Latent Variable Models

There are several event-based variables that enter the latent human rights variable model, which are indicative of information about a specific type of repressive event that is derived from a pool of regularly augmented primary source documents. 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 US was coded as repressive for certain years. The US received these codings for specific reasons: the US 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 U.S. Senator Joseph McCarthy that were also taking place in the early 1950s. Of course, monitors and the

[4] Fariss (2014).
[5] Coppedge et al. (2014); Pemstein et al. (2015)
[6] Fariss (2014, 2015); Hill Jr. and Jones (2014).

media may be more aware of these events because of the high levels of press freedom in the US relative to other countries. This is an example of the challenges in modeling human rights respect, and it is these very challenges that the new latent variable model helps to reveal and address.

Cingranelli and Filippov (2016) claim that the latent human rights variable estimates are not valid before the Cingranelli and Richards 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 (2015), Cingranelli and Filippov (2016) select specific country-year examples (e.g., 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 Cingranelli and Richards human rights data — models that assume equal weighting of human rights indicators and no error (see Schnakenberg and Fariss, 2014, for further details) — also generate cases with unexpected values (e.g., 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 of interest. These models can bring together information that might be biased in certain ways, explore that bias, and even correct for it. However, identifying every source of bias is a challenging and active area of current research. 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. Creating, validating, and extending latent variable models of repression and other concepts is an iterative process that continues to provide new insights into these important and active areas of measurement research.[7]

Counterintuitive or theoretically anomalous estimates from the latent human rights variable model provide new puzzles and measurement opportunities for enhancing the validity of the existing latent variable estimates. The discovery of deviant cases is essential in the field of measurement because such

[7] See for example Armstrong (2009); Barbera (2015); Bond and Messing (2015); Carroll et al. (2009); Caughey and Warshaw (2015); Clinton, Jackman and Rivers (2004); Crabtree and Fariss (2015); Fariss (2014); Fariss and Schnakenberg (2014); Jackman (2008); Martin and Quinn (2002); McMann et al. (2016); Mislevy (1991); Pemstein et al. (2015); Pemstein, Meserve and Melton (2010); Poole (2005); Poole and Rosenthal (1991, 1997); Rosenthal and Voeten (2004); Schnakenberg and Fariss (2014); Treier and Jackman (2008).

cases raise concurrent validity issues that must be explored.[8] 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.[9]

A deviant case is an observation that is coded at a surprising value or outlier along some theoretical concept.[10] Based on the estimates of the latent human rights variable in 1953, the United States is a likely example of such a deviant case because it receives a very low score relative to most other countries in that year and is probabilistically similar to states like China. This example demonstrates why the latent variable modeling approach is so useful. The model provides much more information about the relationship between country-year units and offers principled modeling strategies for incorporating new theoretical concepts into future versions of the latent variable model.

In 1953, the United States earns its lowest value on the latent variable because it is coded as a "1" for two of the events-based variables described in Fariss (2014). The United States is coded as a "1" by the WHPSI (World Handbook of Political and Social Indicators) Political Executions variable because of the execution of Julius and Ethel Rosenberg.[11] It also was coded as a "1" by Rummell's Genocide and Politicide variable because of politically motivated killings that occurred within the United States.[12] 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 when compared to other similar measures included in the latent variable model[13], such that a coding of "1" occurs if evidence exists that the government in question 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 from Rummel might include widespread killing or killings targeted at political opponents or groups not specified in the definition of geno-politicide used by Harff (2003) and massive repressive events used by Harff and Gurr (1988). This focus on gathering information about all government sanctioned killing

[8] Concurrent validity is a type of construct validity designed to assess the ability of the operationalization of a variable to theoretically distinguish between groups which it should be able to make distinctions (Trochim and Donnelly, 2008).

[9] For a discussion see Schnakenberg and Fariss (2014).

[10] Lijphart (1971); Seawright and Gerring (2008).

[11] Taylor and Jodice (1983).

[12] Rummel (1995); Wayman and Tago (2010).

[13] Harff (2003); Harff and Gurr (1988)

events could mean that there is more bias inherent in the Rummel data compared to the other event-based variables if he missed small scale killings in systematically more repressive or less accessible political contexts.[14] However, 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 work to 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 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 for the counts he generates for the United States. However, for this case, Rummel also considers deaths that occurred within the United States. To mitigate the potential threat to validity of the latent variable estimates, binary event-based indicators were included in the latent variable model in place of the raw event counts.[15]

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 in each year. For this reason, Taylor and Jodice (1983), and the authors from the earlier versions of the dataset originally developed by Russett et al. (1964) and expanded by Taylor and Hudson (1972), used both international and regional sources for every case. Nonetheless, Taylor and Jodice (1983) state in their codebook that the systematic under-reporting from particularly repressive countries could lead to biases in the raw event counts. 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 inclusion in the latent variable model.[16]

What is clear from this discussion is 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 potentially 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

[14] Wayman and Tago (2010) compare the datasets developed by Rummel and Harff. One difference they consider is the number of killings threshold for inclusion in the dataset.

[15] I am currently investigating the incorporation of event count data (and the uncertainty inherent in those counts) into an new expansion of the latent variable model (Fariss, 2013).

[16] Fariss (2014).

the issues revealed by exploring deviant cases such as the United States in 1953. Bias can be identified and dealt with. Indeed, there are some biases that the latent variable model already begins to address such as the changing standard of accountability. For example, the Cingranelli and Richards data categorizes Sweden in 2011 and Guatemala in 1983 as both 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. There are other biases that, to date, the latent variable model does not yet account for. The investigation of other sources of potential bias in the monitoring of human rights is an active area of applied research.[17]

Sweden, as the example above suggests, is a useful example of a deviant case that can reveal important new theoretical concepts related to the human rights monitoring and documentation process. In most years over the last decade, Sweden is coded by the CIRI human rights data as "occasionally" engaging in acts of torture and ill treatment against individuals within that country. Eck and Fariss (2016) explore the features of a country that otherwise ranks among the highest values for many other country level indicators.[18] It turns out that Sweden has developed such high levels of institutional transparency that it is very easy for monitoring organizations to access information of the near census of allegations of this type of abuse or any other form of violence.[19] This high level of institutional transparency likely varies across political contexts and, if this variable is systematically related to the level of human rights, will confound the relationship between the data from the coded human rights reports and the estimated latent variable. The failure to incorporate this into future versions of the latent variable model of human rights will continue to bias our cross-sectional comparisons, just as the standard of accountability has biased our temporal comparisons.[20]

As the examples presented in this section illustrate, deviant cases are useful research design tools because they help to identify new theoretical concepts that can inform extensions of the latent variable

[17] Bagozzi and Berliner (2015); Clark and Sikkink (2013); Dancy and Fariss (2017); Eck and Fariss (2016); Fariss et al. (2015); Hill Jr., Moore and Mukherjee (2013).

[18] Schell et al. (2007).

[19] As Eck and Fariss (2016) describe, Sweden has the highest reported rate of rape in Europe (Hofer, 2000). However, institutional transparency is probably not the only reason why this type of violence is reported at such high levels (relative to other social contexts). Instead, it is likely that in Sweden, well developed norms of self-reporting facilitate disclosure of this type of violence. Moreover, the legal definition of rape in Sweden is broader than most other countries. This is similar to how the legal definition of torture and ill-treatment has changed over time based on case law from the European Court of Human Rights (see Fariss, 2014, for additional details).

[20] Fariss (2014, 2015).

model of human rights. In the next section, I discuss additional validity assessments, specifically regarding the increase in respect for human rights in both democratic and non-democratic states. I compare the increasing trends for the latent human rights variable from the changing standard of accountability model with the trends obtained from the latent variable estimates that ignore this concept.

3 Validating Patterns of Human Rights and Democracy

Researchers should focus future efforts on assessing the validity of the latent variable model parameters in addition to the underlying data or information itself. In so doing, applied researchers create new opportunities for creating, extending, and validating measurements of a wide range of concepts. Any of the limitations or flaws inherent in the raw text contained in the human rights documents can be accounted for in the setup of the latent variable model itself. Indeed, most producers of human rights data acknowledge the inherent limitations of the qualitative documentary evidence used to produce and corroborate data. For example, Conrad, Haglund and Moore (2013) state that their data are designed to capture "reporting" and not the census of allegations of ill-treatment and torture. Relatedly, Eck and Hultman (2007) also acknowledge the uncertainty of the estimation process of indicators of political violence by providing a best estimate, and a high and a low estimate of government one-sided killings. Wood and Gibney (2010, 372) instruct the coders of the Political Terror Scale (PTS) "to presume that the information in the [human rights] reports is accurate and complete. Thus, any biases found in the annual reports of the two organizations should be evident in the PTS indices."

One type of test of the validity of an estimated variable involves comparing it to another variable to which it should theoretically relate.[21] One such relationship, that a great deal of scholars have considered, is the relationship between democracy and human rights.[22] Cingranelli and Filippov (2016) graph the polity2 and latent human rights variables together and discuss the visual patterns. However, these authors misinterpret the positive changes in respect for human rights by both democratic and non-democratic

[21] Adcock and Collier (2001) refer to this type of validity as nomological validity or simply construct validity. Trochim and Donnelly (2008) refer to this type of validity as predictive validity. See also Campbell (1960).

[22] See for example Davenport (2010); Davenport and Armstrong (2004); Hill Jr. (2016); Poe and Tate (1994); Poe, Tate and Keith (1999). However, Hill Jr. (2016) and Hill Jr. and Jones (2014) have recently analyzed the conceptual and operational overlap between human rights and democracy. This overlap makes the statistical analysis of these variables in relationship to one another problematic. These authors advise researchers to exercise caution when evaluating the empirical relationship between these two variables.

states that have occurred over time when using the latent variable that accounts for the changing standard of accountability. 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.

That the number and proportion of democratic states is increasing in the international system over time is not contested (see Figure 1). In contrast to this general increase over time however, 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 Figure 2 and Figure 3). 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 when comparing democracies with non-democracies. However, without the assumption of the changing standard of accountability, one must believe that the level of human rights has been steadily decreasing 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.

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.[23] 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. This difference provides evidence of the improvement in predictive (nomological) validity for the latent variable model that accounts for the changing standard of accountability, as recent research has demonstrated that these

[23]The correlation coefficient between polity2 and the CIRI physical integrity index is 0.391 [95% Confidence Interval: 0.365 0.416].

two concepts are strongly related to one another both conceptually and operationally.[24]

However, these correlation coefficients raise an important theoretical issue that complicates regression based analysis of these two variables, which Cingranelli and Filippov (2016) do not consider. Hill Jr. (2016) and Hill Jr. and Jones (2014) analyzed the conceptual and operational overlap between measures of human rights and democracy. These authors argue 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 (2015), this is not the case for every model specification because of the potential bias caused by the operational overlap between the human rights dependent variable and the democracy independent variable. Importantly, "even though the individual [regression model] coefficients change depending on the model specification, the differences are consistent, which is a substantively important finding that eliminates concern that the use of a particular control variable is driving the results. The differences between coefficients are therefore robust to variable selection".[25]

In addition to the selection of variables, the selection of the start year for the sample of country-year units is another important modeling consideration. Most of the treaties analyzed by human rights scholars come into force before the start year of many existing human rights datasets such as the Cingranelli and Richards data (1981-2011). The truncation of the sample to 1981 excludes country-years that had already had the opportunity to ratify available human rights treaties. To justify their truncation decision, Cingranelli and Filippov (2016) suggest the latent variable estimates are not valid before the Cingranelli and Richards human rights data series begins in 1981 because the latent variable estimates are extrapolations "based on very sporadic and eclectic bits of information" (pg.TBD). As already discussed in the previous section, 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[26], a measure of genocide starting in 1956[27], a measure of massive repressive events beginning in 1945[28], a measure of democide/politicide beginning in 1949[29], and a measure of political executions

[24] Hill Jr. (2016); Hill Jr. and Jones (2014).
[25] Fariss (2015, TBD).
[26] Gibney, Cornett and Wood (2012)
[27] Harff (2003)
[28] Harff and Gurr (1988)
[29] Rummel (1995); Wayman and Tago (2010)

beginning in 1948.[30] These are not just "eclectic bits" of data but well documented and relable indicators of repression, which are available for many of the country-year units that enter the model (see Figure 6 for the temporal availability of these data). More to the point, 1965 is the year in which the first human rights treaty is open for signature (International Convention on the Elimination of All Forms of Racial Discrimination). This treaty comes into force in 1969. Thus, this truncation decision complicates the analysis of pooled cross-sectional time series data by removing country-year observations in years prior to the ratification of certain UN human rights treaties. The next section presents replication tests that assess the results using different start years for the regression analyses present in Fariss (2015) in addition to replications that use newly available human rights data the Varieties of Democracy Project.[31]

4 The Association Between Treaty Ratification and Compliance

4.1 Human Rights and Treaty Ratification Replication Analysis

Recall that in Fariss (2015), two linear model coefficients are compared using the dependent variable from the latent variable model that does not account for the changing standard of accountability (labeled the constant standard model) and the dependent variable from the latent variable model that does account for the changing standard of accountability (labeled the dynamic standard model). These competing dependent variables are regressed on 10 treaty variables, including a latent treaty variable, two versions of an additive treaty scale, a proportion of the total number of ratified treaties over the treaties open for ratification, and six binary variables. Each binary treaty variable measures whether or not a country has ratified the Convention Against Torture (CAT), the Convention on the Elimination of all Forms of Discrimination Against Women (CEDW), the Covenant on Civil and Political Rights (CCPR), Covenant on Civil and Political Rights (CCPR), Convention on the Rights of the Child (CRC), or Convention on the Elimination of All Forms of Racial Discrimination (CERD) in a given year.[32]

[30] Taylor and Jodice (1983)

[31] Coppedge et al. (2014); Pemstein, Tzelgov and ting Wang (2015).

[32] Several control variables are also included in the eight different specifications for these models. The control variables include a measure of democracy (Marshall, Jaggers and Gurr, 2013), the natural log of GDP per capita (Gleditsch, 2002), the natural log of population (Gleditsch, 2002), and the lagged value of the latent human rights variable and finally the lagged value of one of the various different treaty variables. The choices of variables for these models does not change the difference in the relationship of treaty ratification and respect for human rights. For more information on model specification choices see Fariss (2015) and the Appendix.

Cingranelli and Filippov (2016) replicated the regression models presented in Fariss (2015) after first removing data from 1965-1980. These authors assert that the latent variable model should only be used for years after 1980 because they have implicitly assumed that the CIRI data are the most valid representations of the underlying theoretical concept of human rights when compared with variables coded in earlier years (CIRI data exist from 1981 through 2011).[33] 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. Reducing the sample size in 1981 as opposed to 1980 or earlier is an arbitrary choice. The arbitrariness of this choice becomes obvious when considering the average level of uncertainty for the country-year latent variable estimates each year.

The level of information 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 for 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. The latent variable model is not an extrapolation because there are observed pieces of information for each unit from which the model generates a country-year estimate for the latent variable. Figure 5 displays box plots for 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. The availability of information over time is displayed in Figure 6. This uncertainty is incorporated into the regression models presented in Fariss (2015).

Cingranelli and Filippov (2016) use the regression estimates based on the reduced sample size to argue that human rights are not associated with treaty ratification because some of these specifications generate insignificant results. However, that there is no evidence for a positive correlation between human rights compliance and treaty ratification for some model specifications is nonetheless important evidence that contradicts the findings from many existing studies, which report negative correlations between these variables. The bulk of the evidence presented in Fariss (2015) and additional evidence presented in this

[33] An evaluation of the validity claims made in the debate between Cingranelli and Richards (2010) and Wood and Gibney (2010) is beyond the scope of this paper.

section supports the inference that there is a positive association between these variables.

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). These regression models are estimated from samples that each contain country-year units starting in each successive year (not just 1965 or 1981). 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 coefficients begin to become statistically indistinguishable from 0 using a p-value threshold of 0.05. Importantly however, the regression coefficients from the two competing models themselves also become statistically indistinguishable from 0. The figures in the appendix represent 62 samples (the start year for each sample increases from 1949 through 2010), for 2 competing dependent variables, for 8 different regression model specifications, for 10 different treaty variables, or $62 * 2 * 8 * 10 = 9,920$ regression models. In the Appendix, interested readers can select their preferred model specification and sample time-frame. However, the eventual lack of statistical significance is not surprising given that the number of units is decreasing, but also because as the start year for the sample increases, a greater number of countries enter the sample having already ratified an increasing number of treaties.[34]

To reiterate, the regression coefficients that represent the relationship between treaty ratification and human rights compliance become statistically indistinguishable from 0 for most of the models once the sample size is reduced by including only years after the early to mid 1980s. Examining the full range of possible specifications shows that, 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. When evaluating the full sample, there is a positive relationship. If we restrict the sample to 1980-2010 or any other start-year thereafter, the model usually generates an insignificant result. Again though, these choices are arbitrary, especially since most of the treaties under study were open for ratification prior to 1981. Nonetheless, the results

[34]Note that these models are not designed for causal inference and, though a variety of selection issues are known to exist when using this model specification (see discussions in Neumayer (2005), Simmons and Hopkins (2005), Von Stein (2005), Simmons (2009), Hill Jr. (2010), and most recently Conrad and Ritter (2013) and Lupu (2013)), the results are still consistently positive, which again casts doubt on earlier research that found negative correlations between human rights compliance and treaty ratification.

reported in Fariss (2014), Fariss (2015), Cingranelli and Filippov (2016), and Hill Jr. and Jones (2014)[35] directly contradict the negative correlations reported in earlier studies[36] and cast considerable doubt on studies that begin with this negative correlation as a puzzle needing to be explained.[37]

4.2 New Human Rights Data and Treaty Ratification Replication Analysis

The best strategy for assessing the validity of an inference is to replicate the test with new data.[38] New, expert-coded human rights indicators were recently published as part of the Varieties of Democracy Project.[39] The expert-coded data makes use of multiple coders per country-year unit (at least 5 coders per unit but often many more) to generate latent scores based on categorical questions answered by each coder for each country-year item.[40] The model even 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.[41] The project has generated more than 300 variables at present spanning the years 1900-2015. 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 for more details).

Importantly, unlike the standards-based human rights data, the V-DEM project controls the standards used to assess each of their variables (i.e., 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. However, like the event-based data, the V-DEM expert-coders are reliant on their knowledge of evidence from

[35] The recent article by Hill Jr. and Jones (2014) offers additional evidence in support of this inference. These authors use cross-validation and random forest methods to determine the predictive power of the covariates identified as important in the literature on human rights using the existing CIRI and PTS physical integrity scales and the new latent human rights variable that adjusts for the changing standard of accountability. The cross-validation and random forest methods corroborate the results that ratification of the Convention Against Torture and the International Covenant on Civil and Political Rights are positively associated with respect for human rights. The authors do not consider any other UN human rights treaty variables.

[36] Hathaway (2002); Hafner-Burton and Tsutsui (2005, 2007)

[37] Hafner-Burton (2013); Hollyer and Rosendorff (2011); Posner (2014); Vreeland (2008).

[38] For a discussion of different types of replication studies see Fariss and Jones (Forthcoming).

[39] Coppedge et al. (2014); Pemstein et al. (2015); Pemstein, Tzelgov and ting Wang (2015).

[40] There are some rare exceptions to this rule. Special thanks to the V-DEM team for alerting me to this detail.

[41] McMann et al. (2016); Pemstein et al. (2015).

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.[42] 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 future research that will hopefully inform new versions of the latent human rights model.

Figure 7 plots the yearly average for the two V-DEM human rights variables from 1949-2013, the same time period available for the updated version of latent human rights variable. What should be clear from this visualization, is a very similar upward trend in human rights respect after the end of Cold War. This upward trend is consistent with the pattern of the latent variable that accounts for the changing standard of accountability in the upper panel of Figure 4. These similar patterns provide evidence of the convergent validity of the latent human rights variable that incorporates the changing standard of accountability. According to the V-DEM human rights data and consistent with previous findings, human rights are improving over time.[43]

Correlation coefficients between the two V-DEM variables and the two versions of the latent human rights variables support this inference.[44] The correlation coefficients for the V-DEM variables and the latent human rights variable that incorporates the changing standard of accountability are larger than the same statistics estimated using the latent human rights variable with a constant standard of accountability. The correlation coefficient between posterior draws of the V-DEM torture variable and posterior draws of the latent human rights variables are 0.632 [95% *Credible Interval* : 0.624, 0.639] (changing standard of accountability) and 0.568 [95% *Credible Interval* : 0.560, 0.575] (constant standard of accountability) respectively. The difference between these estimates is 0.064 [95% *Credible Interval* : 0.054, 0.075]. The correlation coefficient between posterior draws of the V-DEM killing variable and posterior draws of the latent human rights variables are 0.642 [95% *Credible Interval* : 0.635, 0.650] (changing standard of

[42] Lustik (1996).

[43] Fariss (2014).

[44] The correlation coefficients are calculated by taking draws from the country-year posterior distributions for each variable and then calculating the correlation coefficient for each pair of draw.

accountability) and 0.576 [95% *Credible Interval* : 0.569, 0.584] (constant standard of accountability) respectively. The difference between these estimates is 0.066 [95% *Credible Interval* : 0.055, 0.076]. [45]

These results add support to the inferences about improving human rights over time presented in Fariss (2014) and the positive association between human rights compliance and treaty ratification presented in Fariss (2014) and Fariss (2015). Figure 8 presents visual information for the regression coefficients from models that use the V-DEM human rights variables instead of the latent human rights variables. The model presented here contains all the control variables presented in (Fariss, 2015). All the additional model specifications are available in the Appendix. The figures in the Appendix represent 62 samples (the start year for each sample increases from 1949 through 2010), for the 2 new V-DEM dependent variables, for 8 different regression model specifications, for 4 different treaty variables, or $62 * 2 * 8 * 4 = 3,968$ regression models. The results from these models corroborate the positive correlation found between human rights compliance and treaty ratification reported in (Fariss, 2015) and replicated above.

5 Conclusion

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 because of 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".[46] These are

[45] The correlation coefficients between the point estimates for these latent human rights variables and point estimate of the V-DEM torture variable are 0.697 [95% *Confidence Interval* : 0.686, 0.708] 0.627 [95% *Confidence Interval* : 0.614, 0.640] and the V-DEM killing variable are 0.708 [95% *Confidence Interval* : 0.697, 0.7185] 0.635 [95% *Confidence Interval* : 0.623, 0.6482], which demonstrate the same pattern of increased agreement between the V-DEM variables and the latent human rights variable that incorporates the changing standard of accountability.

[46] 2011, 159.

the reasons why the standard of accountability used to produce human rights documents is becoming increasingly stringent over time and why previous studies have found and reported negative patterns instead of positive ones.

Cingranelli and Filippov (2016), raise several issues with the positive relationships drawn from the latent variable model of human rights. I am grateful to these authors for raising these points because they highlight critical issues in the measurement and study of human rights. Their conclusion that respect for human rights is declining and unaffected by treaty ratification, however, is not supported by the available empirical evidence. This essay takes three steps to clarify and improve the measurement of human rights and to address the concerns of Cingranelli and Filippov. First, it outlines a deviant-case study design that provides new insights into improving the existing latent human rights variable. Second, it offers new evidence that the standard of accountability in human rights monitoring is indeed changing over time for democracies and autocracies. And third, it presents new replication analyses, using both old and new human rights data, that support the finding of a positive trend in respect for human rights and a positive association between respect for human rights and treaty ratification.

A science of human rights requires valid comparisons of repression levels across different spatial and temporal contexts.[47] Scholars of human rights, repression, and contentious politics are aware that the quality of information across political contexts is inconsistent.[48] The heterogeneity in the quality and availability of information is a major stumbling block for the generation of valid inferences about topics of interest to our research community. Though there is no model-free way to estimate unobservable concepts such as human rights, latent variable models, with their focus on the theoretical relationship between data and model parameters, offer a principled way to bring together this information, make sense of it, and then using it, discover new theoretical advances in understanding the dynamic patterns of human rights.

[47] Schnakenberg and Fariss (2014).

[48] See for example Brysk (1994); Carleton and Stohl (1985); Conrad, Haglund and Moore (2013); Eck and Fariss (2016); Eck and Hultman (2007); Fariss (2014); Jabine and Claude (1992); Poe, Carey and Vazquez (2001); Schnakenberg and Fariss (2014); Stohl, Carleton and Johnson (1984); Wood and Gibney (2010).

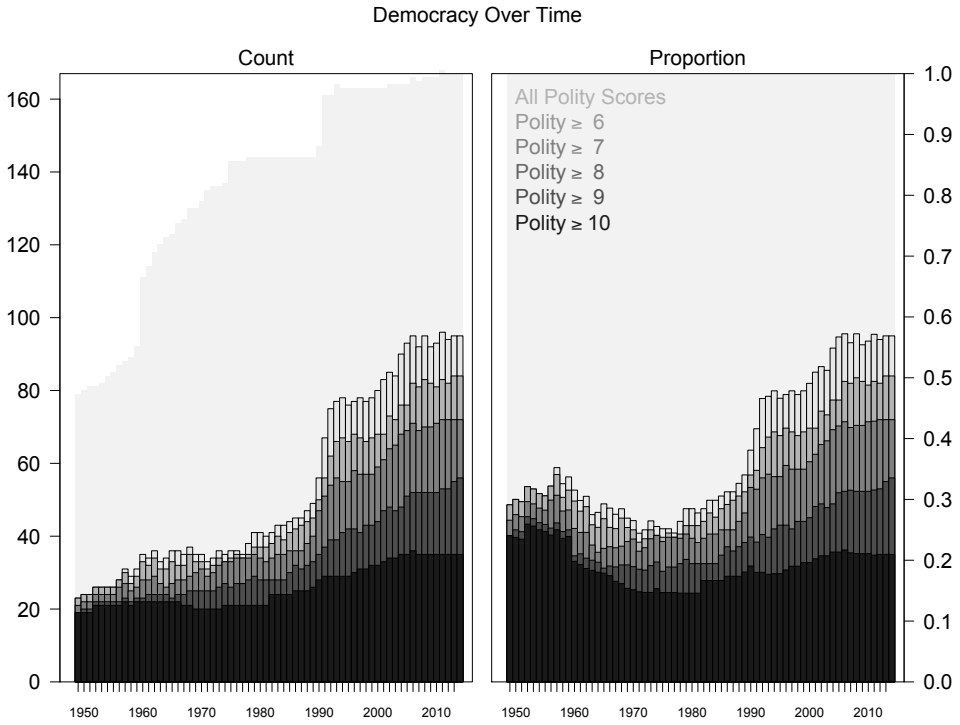

Figure 1: The absolute number of democracies in the international system increases over time though the proportion of democracies in the system has not increased for the highest category on the polity IV scale. The polity IV dataset currently only covers 167 states.

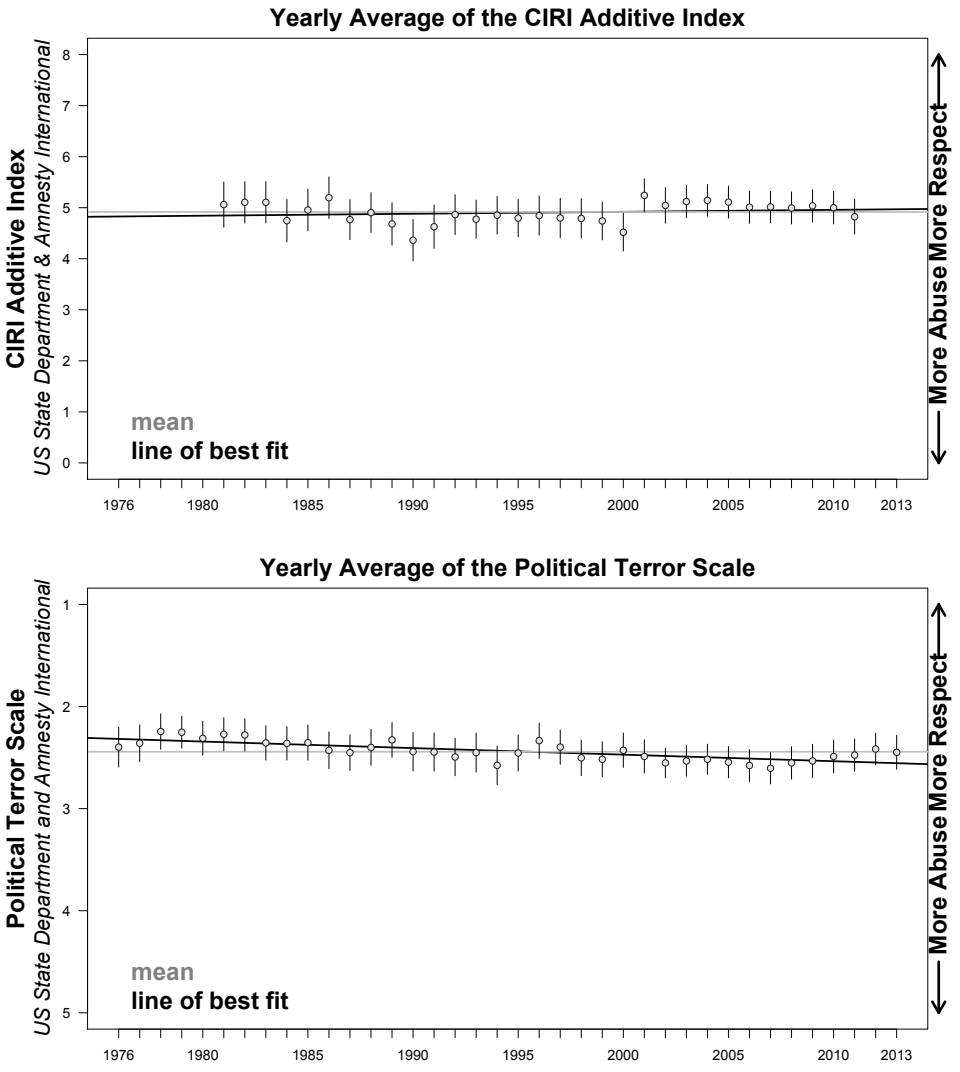

Figure 2: Yearly mean and 95% confidence intervals for the estimated level of repression using the CIRI Additive index (upper panel), and the Political Terror Scale index (lower panel). Each series is based on the human rights reports from the US State Department and Amnesty International. Note that the averages for the Political Terror Scale estimates are based on two scales coded independently, one from the US State Department reports and one from the Amnesty International reports.

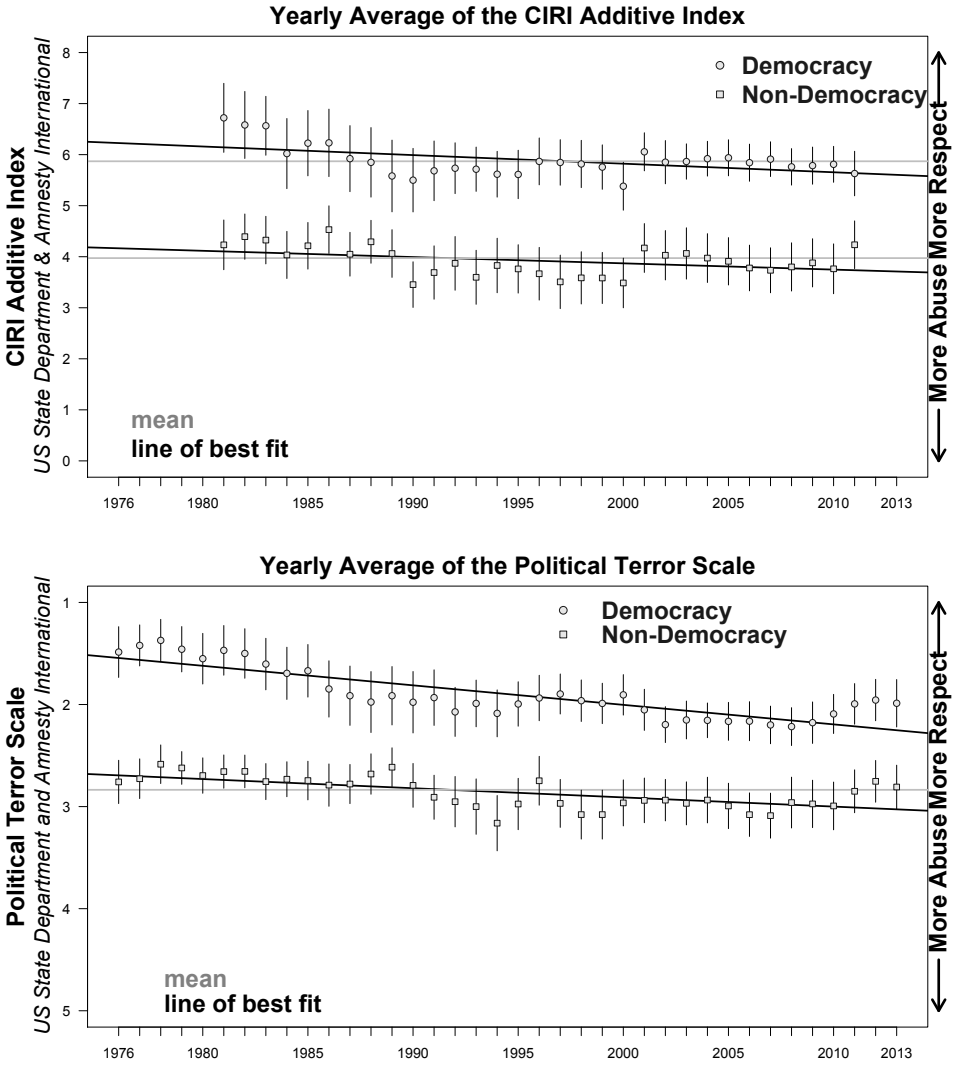

Figure 3: Yearly mean and 95% confidence intervals for the estimated level of repression using the CIRI Additive index (upper panel), and the Political Terror Scale index (lower panel) for democratic and non-democratic states as measured by Polity IV (values of 6 or greater). Each series is based on the human rights reports from the US State Department and Amnesty International. The level of human rights decreases for both democracies and non-democracies according to the CIRI physical integrity index and the Political Terror Scale.

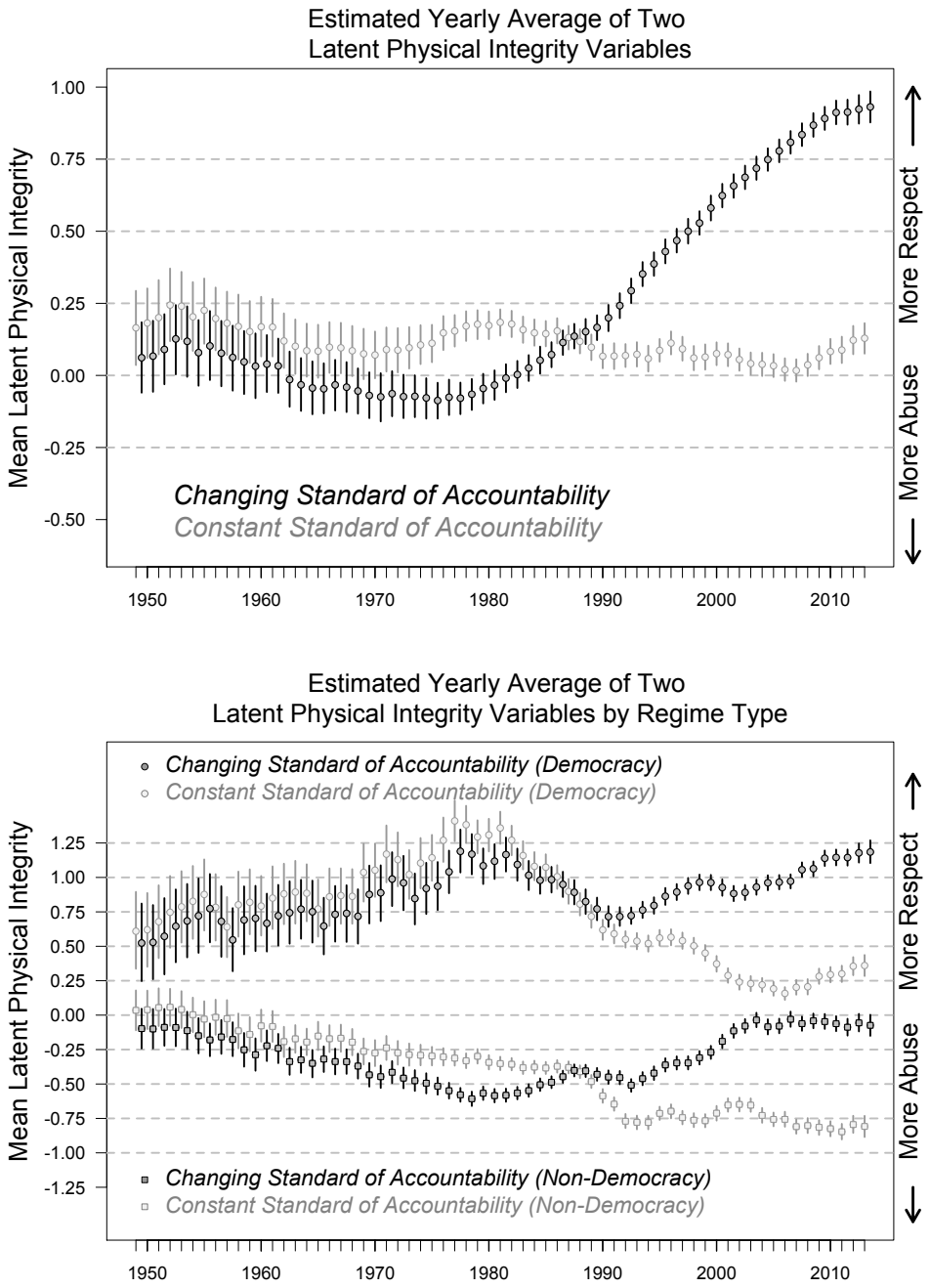

Figure 4: The upper panel displays yearly mean and credible intervals for latent physical integrity estimates from the changing standard latent variable model and the constant standard latent variable model. The lower panel displays yearly mean and credible intervals for these same variables across democratic and non-democratic states as measured by Polity IV (values of 6 or greater). Only the latent variable estimates that assume a changing standard of accountability show improvement for either type of country-year. Without the assumption of the changing standard of accountability, one must believe that the level of human rights in just the set of democratic states has been steadily decreasing since a high point in the early 1980s. It is more likely 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.

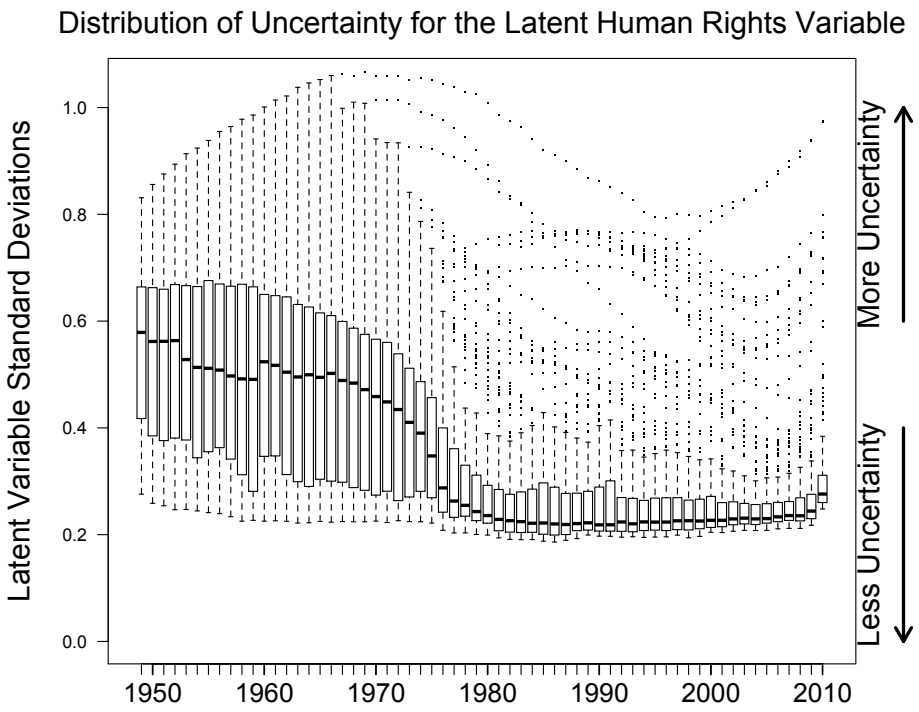

Figure 5: The yearly distribution of the standard deviations from the latent variable estimates from 1949-2010. Though not every one of the repression variables is measured for each country-year unit, the latent variable model is able estimate a value of the latent variable for each country-year unit using the observed variables that are available. As this graph illustrates, the level of uncertainty for each country-year unit is in part a function of the availability of the observed variables. Thus, there is more uncertainty in earlier years and importantly this uncertainty information can be incorporated into standard statistical analyses (Schnakenberg and Fariss, 2014). As new repression variables are incorporated into future versions of the latent human rights model, these estimates will decrease, conditional on the relative quality of those new variables.

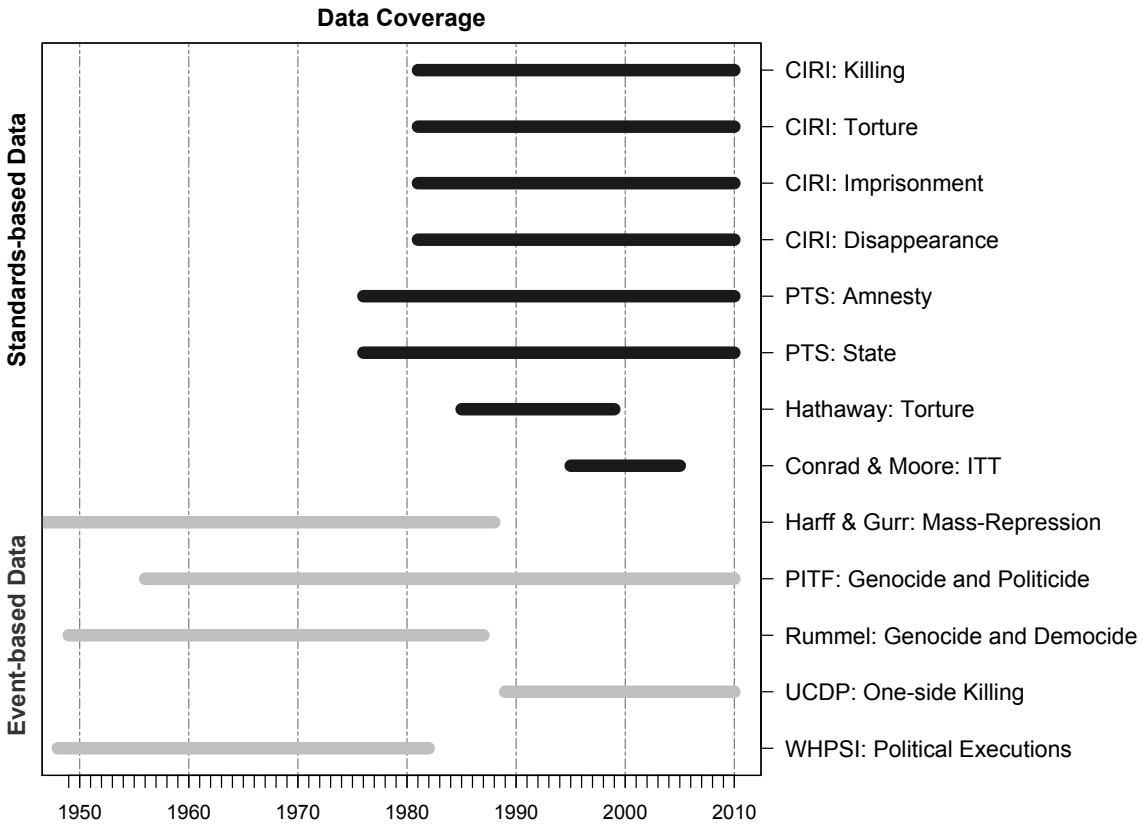

Figure 6: Temporal coverage and data type of repression data sources. Grey lines are event-based data. Black lines are standards-based measures. As additional data becomes available, the level of uncertainty of the latent variable estimates decrease as show in Figure 5.

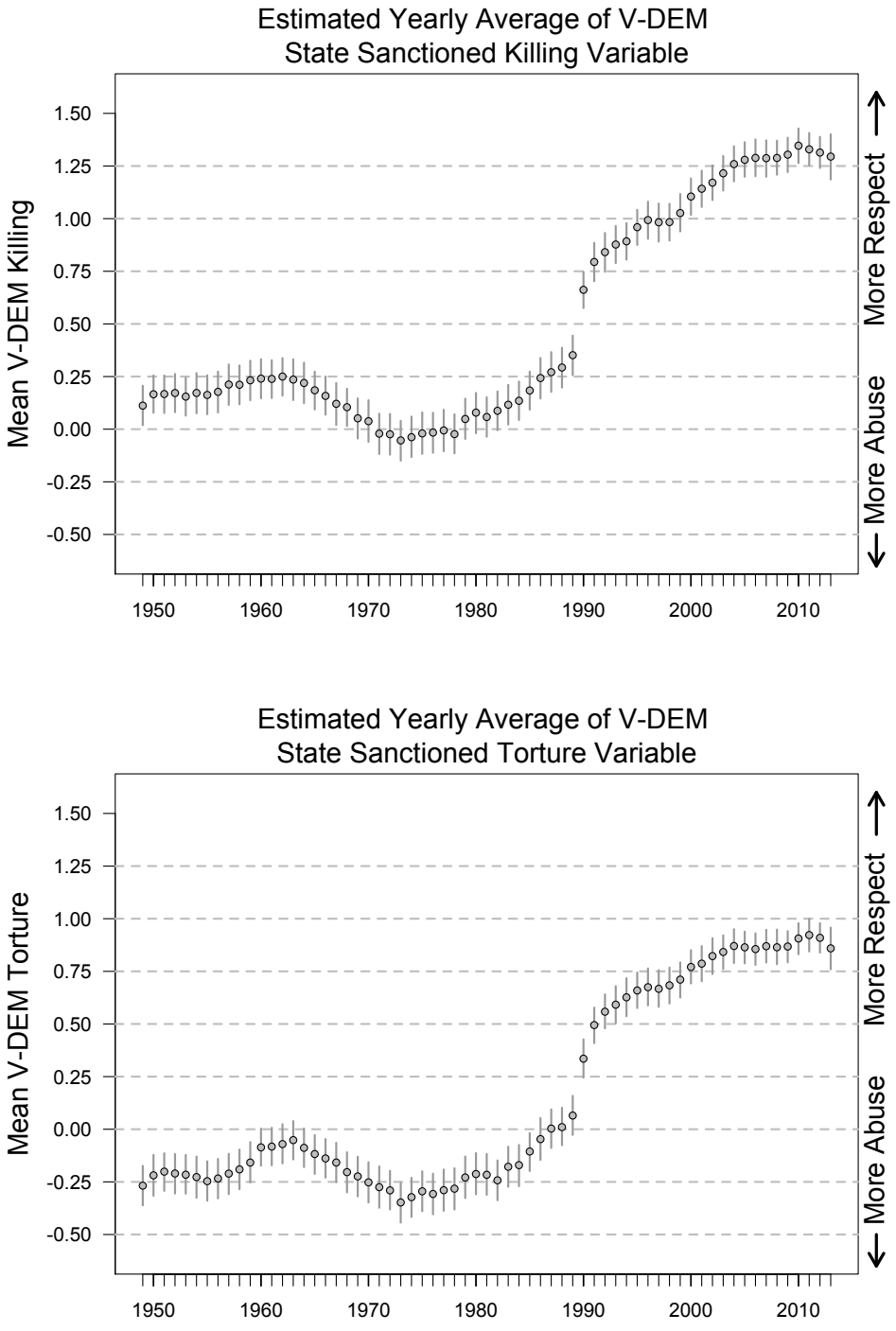

Figure 7: The yearly average for the two expert-coded V-DEM physical integrity variables from 1949-2013 (Coppedge et al., 2014; Pemstein, Tzelgov and ting Wang, 2015), which is the same time period available for the most recent update of latent human rights variable. What should be clear from this visualization, is a very similar upward trend in human rights respect after the end of Cold War. This upward trend is consistent with the pattern of the latent variable that accounts for the changing standard of accountability in the upper panel of Figure 4. These similar patterns provide evidence of the convergent validity of the latent human rights variable that incorporates the changing standard of accountability. According to the V-DEM human rights data and consistent with previous findings (Fariss, 2014), human rights are improving over time.

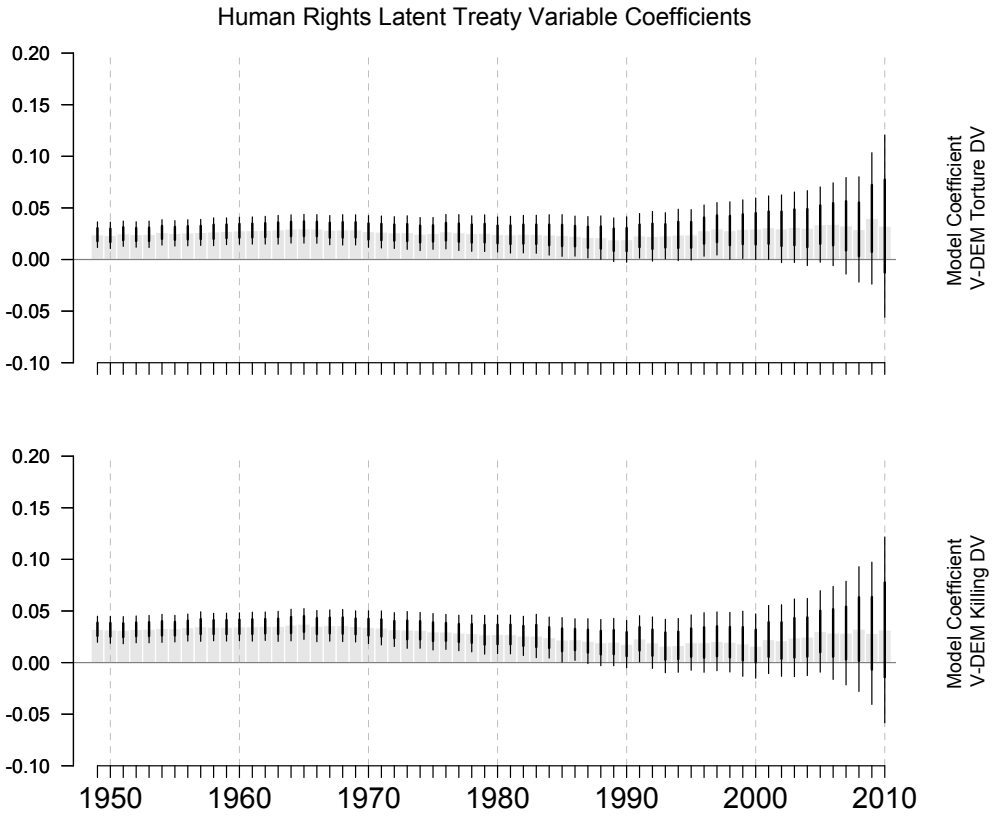

Figure 8: Positive regression coefficients from regression models that regress the two V-DEM human rights variables (Coppedge et al., 2014; Pemstein, Tzelgov and ting Wang, 2015) on the latent treaty variable in addition to the control variables described in Fariss (2015). These model coefficients corroborate results for a positive correlation between human rights compliance and treaty ratification. The regression model specification that produces these coefficients contains all of the control variables considered in Fariss (2015). Each of the 8 model specifications for the two new V-DEM human rights variables are displayed visually in the Appendix.

References

Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3):529–546.

Armstrong, David A. 2009. "Measuring the democracy-repression nexus." *Electoral Studies* 28(3):403–412.

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." unpublished manuscript.

Barbera, Pablo. 2015. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.

Bond, Robert M. and Solomon Messing. 2015. "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook." *American Political Science Review* 109(1):62–78.

Brysk, Alison. 1994. "The Politics of Measurement: The Contested Count of the Disappeared in Argentina." *Human Rights Quarterly* 16(4):676–692.

Campbell, Donald T. 1960. "Recommendation for APA Test Standards Regarding Construct, Trait, or Discriminant Validity." *American Psychologist* 15(August):546–553.

Carleton, David and Michael Stohl. 1985. "The Foreign Policy of Human Rights: Rhetoric and Reality from Jimmy Carter to Ronald Reagan." *Human Rights Quarterly* 7(2):205–229.

Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole and Howard Rosenthal. 2009. "Comparing NOMINATE and IDEAL: Points of Difference and Monte Carlo Tests." *Legislative Studies Quarterly* 34(4):555–591.

Caughey, Devin and Christopher Warshaw. 2015. "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model." *Political Analysis* 23(2):197–211.

Cingranelli, David L. and David L. Richards. 2010. "The Cingranelli and Richards (CIRI) Human Rights Data Project." *Human Rights Quarterly* 32(2):401–424.

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):539–568.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2):355–370.

Conrad, Courtenay R. and Emily Hencken Ritter. 2013. "Treaties, Tenure, and Torture: The Conflicting Domestic Effects of International Law." *Journal of Politics* 75(2):397–409.

Conrad, Courtenay R., Jillienne Haglund and Will H. Moore. 2013. "Disaggregating Torture Allegations: Introducing the Ill-Treatment and Torture (ITT) Country-Year Data." *International Studies Perspectives* 14(2):199–220.

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):159–169.

Crabtree, Charles D. and Christopher J. Fariss. 2015. "Uncovering Patterns Among Latent Variables: Human Rights and De Facto Judicial Independence." *Research & Politics* 2(3):DOI: 10.1177/2053168015605343.

Dancy, Geoff and Christopher J. Fariss. 2017. "Rescuing Human Rights Law from International Legalism and its Critics." *Human Rights Quarterly* Forthcoming.

Davenport, Christian. 2010. *State Repression and the Domestic Democratic Peace*. New York: Cambridge University Press.

Davenport, Christian and David A. Armstrong. 2004. "Democracy and the violation of human rights: A statistical analysis from 1976 to 1996." *American Journal of Political Science* 48(3):538–554.

Eck, Kristine and Christopher J. Fariss. 2016. "Ill Treatment and Torture in Sweden: A Critique of Cross-Case Comparisons.".

Eck, Kristine and Lisa Hultman. 2007. "Violence Against Civilians in War." *Journal of Peace Research* 44(2):233–246.

Fariss, Christopher J. 2013. "Uncertain Events: A Dynamic Latent Variable Model of Human Rights Respect and Government Killing with Binary, Ordered, and Count Outcomes." *Presented at the Annual Meeting of the Society for Political Methodology* .

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.

Fariss, Christopher J. 2015. "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 Keith Schnakenberg. 2014. "Measuring Mutual Dependence Between State Repressive Actions." *Journal of Conflict Resolution* 58(6):1003–1032.

Fariss, Christopher J. and Zachary M. Jones. Forthcoming. "Enhancing Validity in Observational Settings When Replication is Not Possible." *Political Science Research and Methods* .

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

Gleditsch, Kristian Skrede. 2002. "Expanded trade and GDP data." *Journal of Conflict Resolution* 46(5):712–724.

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):1373–1411.

Hafner-Burton, Emilie M. and Kiyoteru Tsutsui. 2007. "Justice lost! The failure of international human rights law to matter where needed most." *Journal of Peace Research* 44(4):407–425.

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):359–371.

Hathaway, Oona A. 2002. "Do human rights treaties make a difference?" *Yale Law Journal* 111(8):1935–2042.

Hill Jr., Daniel W. 2010. "Estimating the Effects of Human Rights Treaties on State Behavior." *Journal of Politics* 72(4):1161–1174.

Hill Jr., Daniel W. 2016. "Democracy and the Concept of Personal Integrity Rights." *Journal of Politics* .

Hill Jr., Daniel W., 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):219–232.

Hill Jr., Daniel W. and Zachary M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108(3):661–687.

Hofer, Hanns Von. 2000. "Crime Statistics as Constructs: The Case of Swedish Rape Statistics." *European Journal on Criminal Policy and Research* 8(1):77–89.

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.

Jabine, Thomas B. and Richard P. Claude, eds. 1992. *Human Rights and Statistics: Getting the Record Straight*. Philadelphia, PA: University of Pennsylvania Press.

Jackman, Simon. 2008. Measurement. In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. Oxford University Press.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65(3):682–693.

Lupu, Yonatan. 2013. "The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects." *American Journal of Political Science* 57(4):912–925.

Lustik, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90(3):605–618.

Marshall, Monty, Keith Jaggers and Ted R. Gurr. 2013. "Polity IV Project: Political Regime Characteristics and Transitions 1800-2010 Dataset Users' Manual.".
URL: *www.systemicpeace.org/polity/polity4.htm*

Martin, Andrew D. and Keven M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10(2):134–153.

McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell, and Staffan I. Lindberg. 2016. "Strategies of Validation: Assessing the Varieties of Democracy Corruption Data." *Varieties of Democacy Institute Working Paper 23* pp. 1–45.

Mislevy, Robert. 1991. "Randomization-based inference about latent variables from complex samples." *Psychometrika* 56(2):177–196.

Neumayer, Eric. 2005. "Do International Human Rights Treaties Improve Respect for Human Rights?" *Journal of Conflict Resolution* 49(6):925–953.

Pemstein, Daniel, Eitan Tzelgov and Yi ting Wang. 2015. "Evaluating and Improving Item Response Theory Models for Cross-National Expert Sur." *Varieties of Democacy Institute Working Paper 1* pp. 1–53.

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." *Varieties of Democacy Institute Working Paper 21* pp. 1–27.

Pemstein, Daniel, Stephen A. Meserve and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis to Ten Measure of Regime Type." *Political Analysis* 18(4):426–449.

Poe, Steven C. and C. Neal Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88(4):853–872.

Poe, Steven C., C. Neal Tate and Linda Camp Keith. 1999. "Repression of the Human Right to Personal Integrity Revisited: A Global Cross-National Study Covering the Years 1976-1993." *International Studies Quarterly* 43(2):291–313.

Poe, Steven C., Sabine C. Carey and Tanya C. Vazquez. 2001. "How are These Pictures Different? A Quantitative Comparison of the US State Department and Amnesty International Human Rights Reports, 1976–1995." *Human Rights Quarterly* 23(3):650–677.

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting*. Cambridge: Cambridge University Press.

Poole, Keith T. and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35(1):228–278.

Poole, Keith T. and Howard Rosenthal. 1997. *A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

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

Rosenthal, Howard and Erik Voeten. 2004. "Analyzing roll calls with perfect spatial voting: France 1946-1958." *American Journal of Political Science* 48(3):620–632.

Rummel, Rudolph J. 1995. "Democracy, power, genocide, and mass murder." *Journal of Conflict Resolution* 39(1):3–26.

Russett, Bruce M., Hayward R. Alker Jr., Karl W. Deutsch and Harold D. Lasswell. 1964. *World Handbook of Political and Social Indicators*. New Haven: Yale University Press.

Schell, Carl Otto, Marie Reilly, Hans Rosling, Stefan Peterson and Anna Mia Ekström. 2007. "Socioeconomic determinants of infant mortality: A worldwide study of 152 low-, middle-, and high-income countries." *Scandinavian Journal of Public Health* pp. 288–297.

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*. The Norton Series in World Politics.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge University Press.

Simmons, Beth A. and Daniel J. Hopkins. 2005. "The constraining power of international treaties: Theory and methods." *American Political Science Review* 99(4):623–631.

Stohl, Michael, David Carleton and Steven E. Johnson. 1984. "Human Rights and US Foreign Assistance from Nixon to Carter." *Journal of Peace Research* 21(3):215–226.

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: Yale University Press.

Taylor, Charles Lewis and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators, Second Edition*. New Haven: Yale University Press.

Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.

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

Von Stein, Jana. 2005. "Do treaties constrain or screen? Selection bias and treaty compliance." *American Political Science Review* 99(4):611–622.

Vreeland, James Raymond. 2008. "Political institutions and human rights: Why dictatorships enter into the United Nations Convention Against Torture." *International Organization* 62(1):65–101.

Wayman, Frank W. and Atsushi Tago. 2010. "Explaining the onset of mass killing, 1949–87." *Journal of Peace Research* 47(1):3–13.

Wood, Reed M. and Mark Gibney. 2010. "The Political Terror Scale (PTS): A Re-introduction and Comparison." *Human Rights Quarterly* 32(2):367–400.