Appendix:
Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability

Christopher J. Fariss*

Contents

A Standards-Based Repression Variables 3

B Event-Based Repression Variables 9

C Human Rights Report Example Text 11

D Additional Figures 13

E Country Example Plots 14

F The Changing Standard of Accountability: Additional Plots and Tables 17

G Dynamic Standard Model Cut-points for the Standards-Based Response Variables 30

H Item Discrimination Paramters 41

I Deviance Information Criterion (DIC) 42

J Posterior Predictive Checks: Additional Table 43

K Analyzing the Standards-Based Repression Variables 44

L Analyzing the Event-Based Repression Variables 48

M Method for Incorporating Uncertainty into a Model 50

N Information About The Convention Against Torture Model Specifications 51

O Trace Plots 54

*Pennsylvania State University, Department of Political Science, cjf20@psu.edu; cjf0006@gmail.com

P Latent Variable Model Extensions **55**

Q Bugs Model Files **56**

Introduction to the Appendix

The supplementary material presented in this document provides additional details about the latent variable model developed in the article "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability". The main article makes reference to the materials contained here. The estimates presented in this appendix along with the code necessary to implement the Bayesian models in JAGS will be made publicly available here:

`http://dvn.iq.harvard.edu/dvn/dv/HumanRightsScores`.

A Standards-Based Repression Variables

The standards-based variables that enter the models are derived from Amnesty International and US State Department reports (Cingranelli and Richards, 1999, 2012*a*,*b*; Gibney, Cornett and Wood, 2012; Hathaway, 2002). Another dataset uses "Urgent Action Reports" published throughout the year by Amnesty International to create their index of country-year torture (Conrad and Moore, 2011; Conrad, Haglund and Moore, 2013). I treat this variable as standards-based because the operationalization is based on reports created in a specific historical context just like the other standards-based variables.

Standards-based variables were developed in part because of the availability and comprehensiveness of the human rights reports but also in reaction to the use of event-based data.[1] The Political Terror Scale (PTS) was originally coded by Carleton and Stohl (1985), Gibney and Stohl (1988), Gibney and Dalton (1996) and is now made available by Gibney, Cornett and Wood (2012). The PTS data are two standards-based, 5-point ordinal scales that are respectively measured from the content of the country reports published annually by the US State Department and Amnesty International respectively. Category 1 identifies countries under a secure rule of law, where physical integrity violations like imprisonment, torture, murder and execution do not occur. Countries placed in category 5 are those in which such abuses are a common part of life, affecting all segments of the population. The remaining categories, 2 through 4, represent varying degrees between these two extremes.[2] Some scholars argued that event-based data

[1] Poe (2004) reviews this debate but 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).

[2] The full wording of the PTS coding is below. It is taken from Gastil (1980). See also Gibney and Dalton (1996), Poe and

rather than the standards-based PTS were the more appropriate operationalization of human rights respect (e.g., Davenport, 1995; Lopez and Stohl, 1992).

The CIRI human rights variables are in part an attempt to find a middle ground between the event-based and standards-based data.[3] The variables are still based exclusively on content from the human rights reports but they disaggregate the coding of the four physical integrity rights and use available count information to assess the frequency of violations. Each CIRI human rights variable measures the level of violation on an ordinal scale where, 2 indicates that the right is not violated, 1 indicates that the right is violated occasionally and 0 indicates that the right is violated frequently. Notice that the high values of the CIRI variables measure the highest level of respect for a specific right, whereas the lowest value on the two PTS indices capture the highest level of respect. In many applications that analyze the PTS and CIRI data, one of the indices is usually recoded so that they both measure repression in the same direction. I account for these differences in the model specifications in order to discuss the variables in their original operationalizations. The other two standards-based indices that I discuss next are coded in the same direction as PTS and the opposite direction of the CIRI variables.

Hathaway (2002) and Conrad and Moore (2011) developed indices of country-year torture, using a similar coding scheme. Conrad and Moore (2011) are quick to point out however, that their data are designed to capture "reporting" of torture and not actual "levels" of torture. This is the only dataset with this theoretical distinction. Unlike either the PTS or CIRI variables, the Hathaway (2002) data relies exclusively on content from the US State Department reports to create a 5-point ordinal scale, in which the first level indicates that the reports contained no allegations of torture in a given country-year and level 5 indicates that torture was "prevalent" or "widespread". Levels 2 - 4 represent gradations between these two extremes and again use specific words to map the frequency of torture to a specific level of the variable. The Ill-treatment and Torture (ITT) data use a very similar coding scheme (ranges from 0 to 5). The ITT data are based exclusively on content from "Urgent Action Reports" that are published throughout the year by Amnesty International. In the next subsections, I present the coding rules of the standards-based variables.

Sirirangsi (1993), and Wood and Gibney (2010) for additional discussion of the development of these two indices.

[3]The development of the CIRI data was also in response to criticism of the unidimensionality assumption leveled at PTS by McCormick and Mitchell (1997). By disaggregating the four physical integrity rights into four separate indices, Cingranelli and Richards (1999) demonstrated that the four constructs scaled together into a single unidimensional trait using Mokken Scaling Analysis (Mokken, 1971).

CIRI Physical Integrity Variables (1981-2010)

Each CIRI human rights variable measures the level of violation on an ordinal scale where, 2 indicates that the right is not violated, 1 indicates that the right is violated occasionally and 0 indicates that the right is violated frequently. Notice that the high values of the CIRI variables measure the highest level of respect for a specific right, whereas the lowest value on the two PTS indices capture a highest level of respect. The following descriptions of the four individual physical integrity variables and the physical integrity scale are taken directly from the (Cingranelli and Richards, 2012*a*) code book and discussed at length in (Cingranelli and Richards, 1999):

Extrajudical Killing The variable measuring political and other extrajudicial killings/arbitrary or unlawful depravation of life is coded as a 0 when this practice has occurred frequently in a given year; a score of 1 indicates that extrajudicial killings were practiced occasionally; and a score of 2 indicates that such killings did not occur in a given year.

Disappearance The variable measuring disappearance is coded as a 0 when this practice has occurred frequently in a given year; a score of 1 indicates that disappearances occasionally occurred; and a score of 2 indicates that disappearances did not occur in a given year.

Torture The variable measuring torture and other cruel, inhumane, or degrading treatment or punishment is as coded as a 0 when this practice occurred frequently in a given year; a score of 1 indicates that torture was practiced occasionally; and a score of 2 indicates that torture did not occur in a given year.

Political Imprisonment The variable measuring political imprisonment is coded as a 0 when many people were imprisoned because of religious, political, or other beliefs in a given year; a score of 1 indicates that a few people were imprisoned; and a score of 2 indicates that no persons were imprisoned for any of the above reasons in a given year.

The CIRI coding rules attempt to use count based metrics to rate each of the variables on one of the 3 levels (0, 1, and 2). If the reports provide a count for the number of individuals affected by a given rights violation then following cut offs are used:

Level 0: 50 or more occurrences

Level 1 : From 1 to 49 occurrences

Level 2: Zero occurrences

According to the coder guidelines if an estimate of the number of violations is not be available then the following guidelines from the CIRI code book (Cingranelli and Richards, 2012*a*) are used:

- Instances where violations are described by adjectives such as "gross," "widespread," "systematic," "epidemic," "extensive," "wholesale," routine, regularly, or likewise, are to be coded as a ZERO (have occurred frequently).
- In instances where violations are described by adjectives such as "numerous," "many," "various," or likewise, you will have to use your best judgment from reading through the report to decide whether to assign that country a ONE (have occurred occasionally) or a ZERO (have occurred frequently). Look for language indicating a pattern of abuses; often, these cases merit a ZERO.

Hathaway Torture Scale Coding (1985-1999)

Hathaway (2002) creates a 5-point ordered scale for torture violations. Unlike either the PTS or CIRI variables, the Hathaway (2002) data relies exclusively on content from the US State Department reports. The reports are coded as follows:

- *Level 1*: There are no allegations or instances of torture in this year. There are no allegations or instances of beatings in this year; or there are only isolated reports of beatings by individual police officers or guards all of whom were disciplined when caught.
- *Level 2*: At least one of the following is true: There are only unsubstantiated and likely untrue allegations of torture; there are "isolated" instances of torture for which the government has provided redress; there are allegations or indications of beatings, mistreatment or harsh/rough treatment; there are some incidents of abuse of prisoners or detainees; or abuse or rough treatment occurs "sometimes" or "occasionally." Any reported beatings put a country into at least this category regardless of government systems in place to provide redress (except in the limited circumstances noted above).
- *Level 3*: At least one of the following is true: There are "some" or "occasional" allegations or incidents of torture (even "isolated" incidents unless they have been redressed or are unsubstantiated (see above)); there are "reports," "allegations," or "cases" of torture without reference to frequency; beatings are "common" (or "not uncommon"); there are "isolated" incidents of beatings to death or summary executions (this includes unexplained deaths suspected to be attributed to brutality) or there are beatings to death or summary executions without reference to frequency; there is severe maltreatment of prisoners; there are "numerous" reports of beatings; persons are "often" subjected to beatings; there is "regular" brutality; or psychological punishment is used.
- *Level 4*: At least one of the following is true: Torture is "common"; there are "several" reports of torture; there are "many" or "numerous" allegations of torture; torture is "practiced" (without reference to frequency); there is government apathy or ineffective prevention of torture; psychological

punishment is "frequently" or "often" used; there are "frequent" beatings or rough handling; mistreatment or beating is "routine"; there are "some" or "occasional" incidents of beatings to death; or there are " several" reports of beatings to death.

- *Level 5*: At least one of the following is true: Torture is "prevalent" or "widespread"; there is "repeated" and "methodical" torture; there are "many" incidents of torture; torture is "routine" or standard practice; torture is "frequent"; there are "common," "frequent," or "many" beatings to death or summary executions; or there are "widespread" beatings to death (Hathaway, 2002).

ITT Level of Torture (1995-2005)

Conrad and Moore (2011) have recently released the Ill-treatment and Torture (ITT) project codes the Level of Torture (LoT) using a similar ordinal scale as the ordinal scale developed by (Hathaway, 2002). The variable measures the intensity of government ill-treatment and torture as reported by Amnesty International urgent action reports . The variable captures country-wide allegations of torture that occur throughout the year that used one of the following key words obtained from the documents below. See also the additional discussion of this data by Conrad, Haglund and Moore (2013).

- *Level 0*: None
- *Level 1*: Infrequent
- *Level 2*: Some(times)
- *Level 3*: Frequent
- *Level 4*: Widespread
- *Level 5*: Systematic

The variable measures Amnesty International allegations of the frequency of violations of the practices prohibited by the Convention of Torture throughout a given country during a particular year. Country-year observations with no allegations are coded 0.

Political Terror Scale Coding (1976-2010)

The Political Terror Scale (PTS) was originally coded by Carleton and Stohl (1985), Gibney and Stohl (1988), Gibney and Dalton (1996) and is now made available by Gibney, Cornett and Wood (2012). The PTS data are two standards-based, 5-point ordinal scales that are respectively measured from the

content of the country reports published annually by the US State Department and Amnesty International respectively. See Gibney and Dalton (1996), Poe and Sirirangsi (1993), and Wood and Gibney (2010) for additional discussion of the development of these two indices.

- *Level 1*: Countries under a secure rule of law, people are not imprisoned for their view, and torture is rare or exceptional. Political murders are extremely rare.
- *Level 2*: There is a limited amount of imprisonment for nonviolent political activity. However, few persons are affected, torture and beatings are exceptional. Political murder is rare.
- *Level 3*: There is extensive political imprisonment, or a recent history of such imprisonment. Execution or other political murders and brutality may be common. Unlimited detention, with or without a trial, for political views is accepted.
- *Level 4*: The practices of level 3 are expanded to larger numbers. Murders, disappearances, and torture are a common part of life. In spite of its generality, on this level terror affects those who interest themselves in politics or ideas.
- *Level 5*: The terrors of level 4 have been expanded to the whole population. The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals (Gastil, 1980).

B Event-Based Repression Variables

The event-based binary are variables drawn from Harff and Gurr (1988), Harff (2003), Rummel (1994, 1995), Eck and Hultman (2007), Taylor and Jodice (1983). All of the variables described here are coded 1 if an event described by the different data sources occurred and 0 otherwise.

The early data collection efforts by Rummel and his co-authors predates all other data used in this article (Rummel, 1966, 1994, 1995; Rummel and Tanter, 1974).[4] Davenport (1995), Davenport (1997), and Davis and Ward (1990) used a subset of the indicators collected by Taylor and Jodice (1983). These data have unfortunately fallen out of fashion because they are no longer updated, although there are still examples of recent publications that use this data (Enterline and Gleditsch, 2000; Wayman and Tago, 2010). Harff and Gurr (1988) published data on "massive-repressive" events that included genocide and politicide in addition to large scale repression events. This data by definition captures a larger number of cases than the genocide and politicide data published later by Harff (2003). About half the cases presented in the Harff and Gurr (1988) data are not found in the overlapping time period in the data presented by Harff (2003).

The data produced by Rummel are even more expansive than the other genocide datasets because the data captures "democide", which is defined as killing by government. This broader category is analogous to the one-sided government killing definition that focuses on government caused deaths of non-combatants (Eck and Hultman, 2007). Recently, Wayman and Tago (2010) conducted a thorough comparison of the datasets published by Harff and Gurr (1988), Harff (2003), and Rummel (1994, 1995). Wayman and Tago (2010) caution readers that the existence of these definitional differences need to be considered when comparing results across these data sources. The differences in these definitions are advantageous because each variable is designed to measure the most extreme repressive events but capture some events that do not meet the strictest definition of genocide and politicide.[5] Several of the data sources that publish these binary variables also provide approximations of the number of events that oc-

[4] The methods employed in this article also relate to this early work by Rummel, who was one of the first to adopt factor analysis to model the dimensions of domestic and international violence (Rummel, 1967).

[5] The UN Convention on the Prevention and Punishment of the Crime of Genocide states that Genocide includes only "acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group", which is why analysts include the additional term of politicide. See Ratner and Abrams (2001) for a discussion of the legal developments of this definition.

curred. In another project, I am working on incorporating this information and the uncertainty inherent in estimations of event counts that states have a strategic incentive to hide.[6]

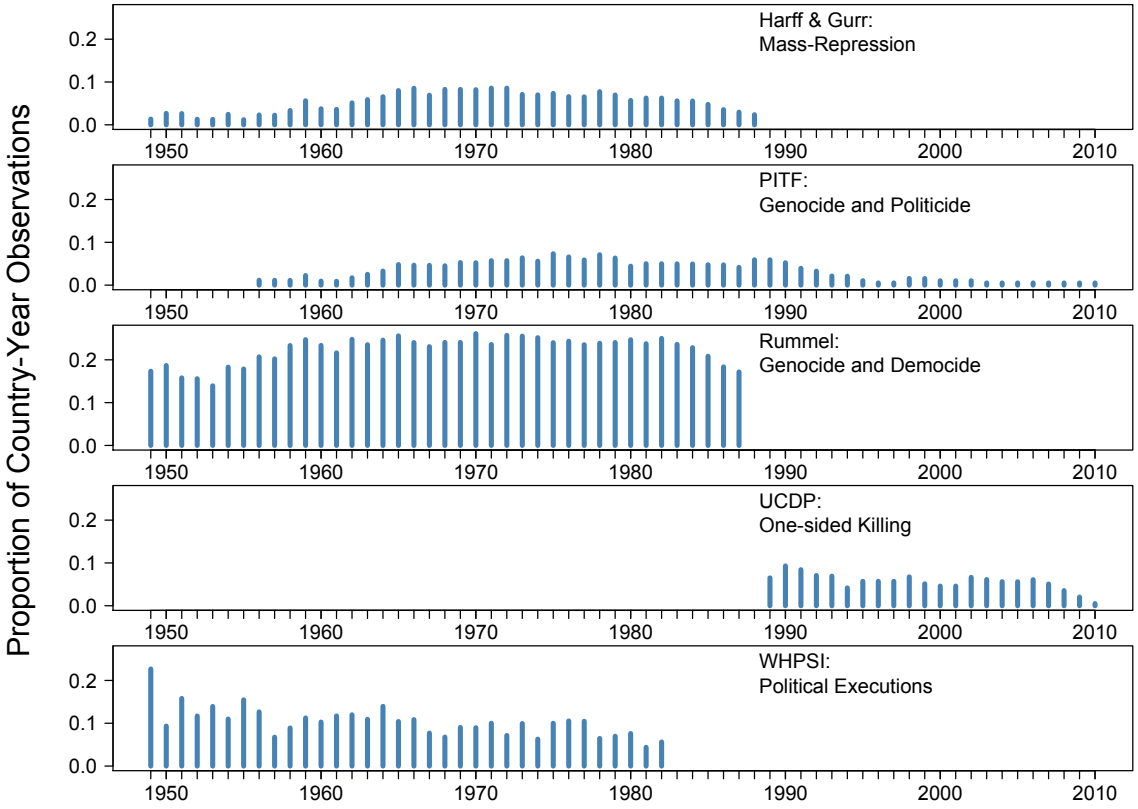

Figure 1:

[6]For a discussion of this issue in the context of counting disappearances see Brysk (1994). See Berman and Clark (1982) and also Clark (2001) for a discussion of assessing the use of disappearances in the context of other rights violations.

C Human Rights Report Example Text

Here I present three examples of text from torture section contained in the State Department human rights reports on Guatemala from the 1981, 1991, and 2001. I also present some of this information in the text of the main article. Note that the example text provides more detailed information in later years and that the raw length of text increases dramatically both for the entire report and the torture section itself as displayed in Table 1. The differences between the coding of "frequent torture" on the CIRI Torture scale in 2001 relative to the less severe coding in 1991 could be a function of the amount of information and the specificity of the information included in the reports in the different years. As these examples suggest, the standard of accountability becomes more stringent as the US State Department and Amnesty International look harder for abuse, look in more places for abuse, and classify more acts as abuse. The reports published today represent a broader and more detailed view of the human rights practices than reports published in previous years. I am exploring the differences in the quality and quantity of information in the text of human rights documents in a book length project that builds off the insights from this article.

Year	Torture Section word count	Full Document word count	CIRI Torture Coding
1981	329	3,930	0 (frequent)
1991	562	5,768	1 (some)
2001	3,669	32,064	0 (frequent)

Table 1: Changing information content in three human rights reports.

Guatemala 1981

> "... the Guatemalan press frequently reports discoveries of bodies evidencing torture. In most instances it has not been possible to establish who the perpetrators were. In some cases there is evidence to suggest that elements within the military or security forces have been involved. In recent months, similar evidence suggests that the guerrilla groups have used torture. ... "

Guatemala 1991

" ... many bodies found throughout Guatemala bore signs of torture or postmortem mutilation. Such treatment, however, is not necessarily evidence of security force involvement: gangs and other criminals, as well as guerrillas, all use torture. There were, nevertheless, many credible reports of torture and mistreatment by security forces. There were also credible reports of the use of excessive force by police at the time of arrest and of abusive treatment by army personnel, civil defense patrols, military commissioners, and police of persons in rural areas. ... "

Guatemala 2001

" ... there were credible reports of torture, abuse, and other mistreatment by members of the PNC during the year. These complaints typically involved the use of excessive force during arrests, interrogations, or other police operations. Criminal Investigative Service (SIC) detectives continued to torture and beat detainees during interrogation to obtain forced confessions. The Government and the PNC showed decreased willingness to investigate, prosecute, or otherwise punish officers who committed abuses. The PNC transferred some cases of alleged torture to the Prosecutor's Office. There were a significant number of murder victims whose bodies demonstrated signs of torture or cruel treatment ... "

D Additional Figures

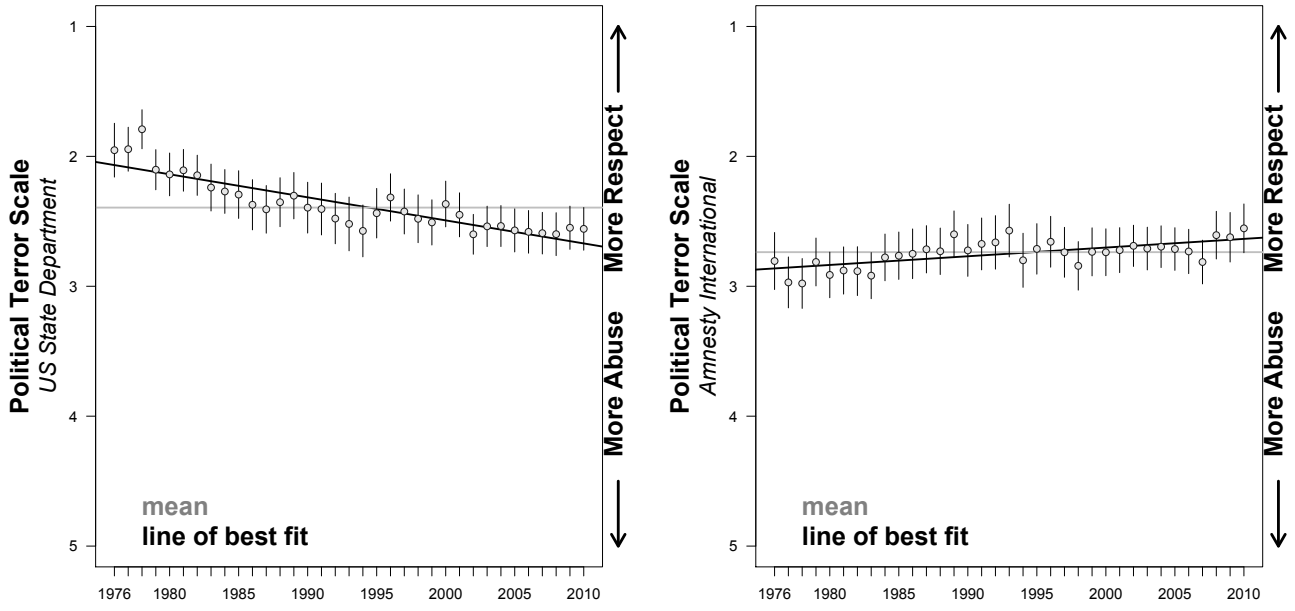

Figure 2: Yearly mean and 95% confidence intervals for the estimated level of repression using the Political Terror Scale index based on the US State Department reports (left), the Political Terror Scale index based on Amnesty International reports (right). Notice that the scales are inverted to be consistent with other figures. The figures suggest individually that the level of repression has changed modestly over time. For the Political Terror Scale estimates based on the State Department reports, the level of repression (respect for rights) has increased (decreased) from 1.952 in 1976 to 2.558 in 2010, a difference of -0.606 [95%:-0.872, -0.340]. However, the opposite trend is observed for the Political Terror Scale index based on the reports from Amnesty International. Here the level of repression (respect for rights) has decreased (increased) from 2.806 in 1976 to 2.554 in 2010, a difference of 0.251 [95%:-0.039, 0.542]. In both cases, the changes are modest but of more substantive importance the changes contradict one another.

E Country Example Plots

Selected country-year posterior estimates and credible intervals (1949-2010). Coverage extends back to 1949 because of the incorporation of multiple indicators of physical integrity rights violations. More information is available about state behavior in the post 1975 period so the estimates are generally more precise from this period onwards. However, the level of precision (inverse variance) is quantified which makes possible probabilistic comparisons across the entire period. The model does a better job of discriminating among abusive states than with states that exhibit moderate to low abuse during the earlier period.

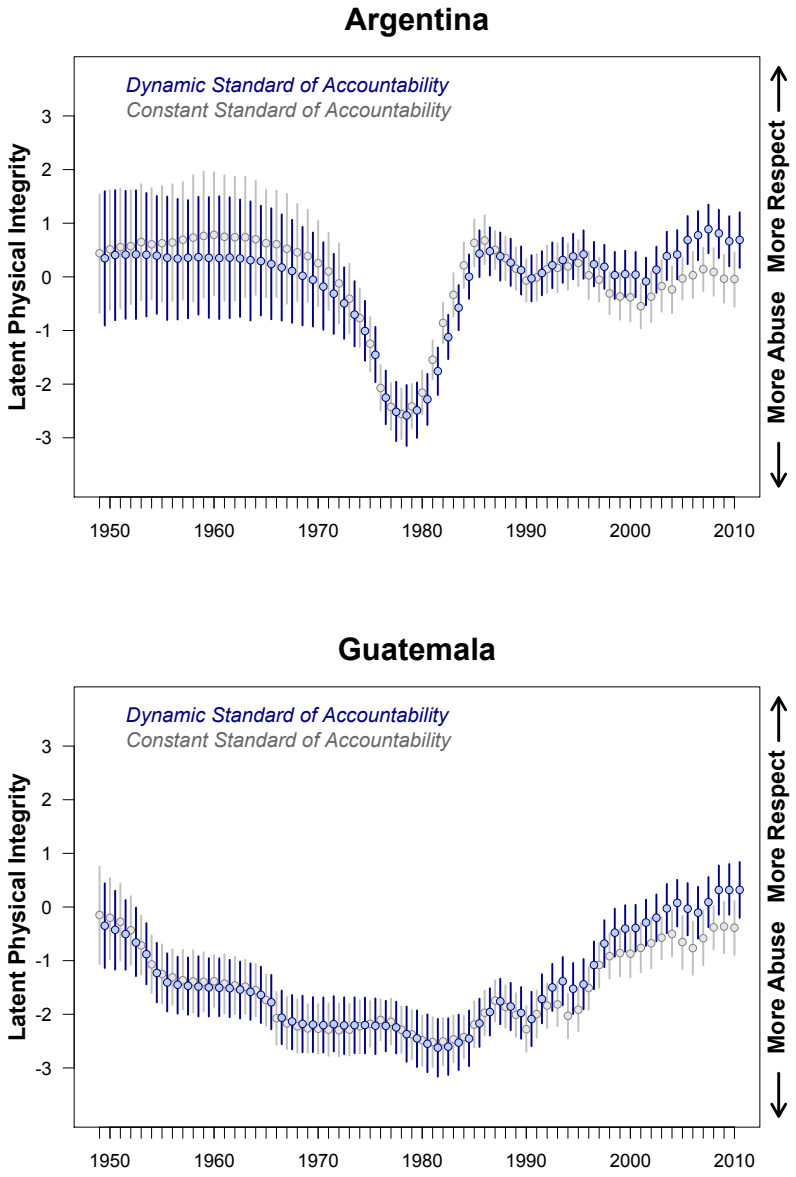

Figure 3: Selected country-year posterior estimates and credible intervals (1949-2010). Coverage extends back to 1949 because of the incorporation of multiple indicators of physical integrity rights violations. More information is available about state behavior in the post 1975 period so the estimates are generally more precise from this period onwards. However, the level of precision (inverse variance) is quantified which makes possible probabilistic comparisons across the entire period. The model does a better job of discriminating among abusive states than with states that exhibit moderate to low abuse during the earlier period. The grey estimates represent those taken from the constant standard model. The blue estimates represent those taken from the dynamic standard model. The dynamic standard model explicitly accounts for changes in the standard of accountability over time. The difference between the two series increases as a function of time.

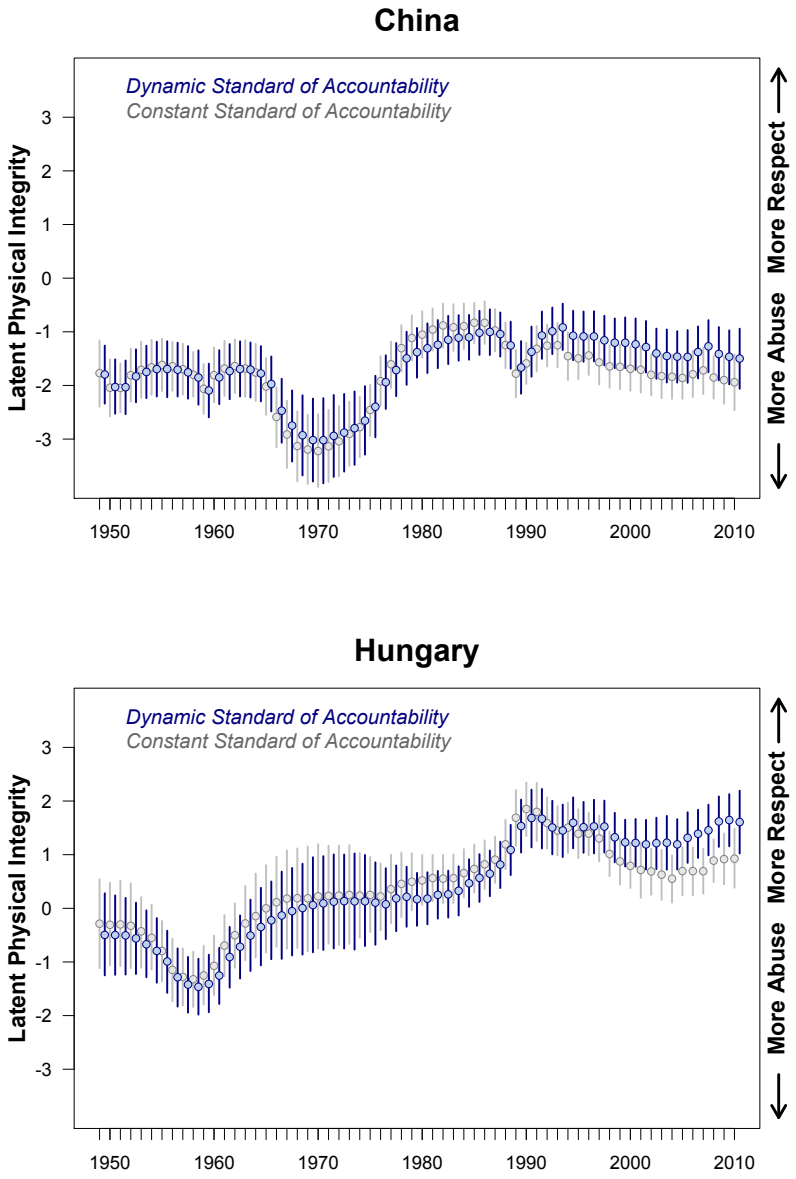

Figure 4: Selected country-year posterior estimates and credible intervals (1949-2010). The grey estimates represent those taken from the constant standard model. The blue estimates represent those taken from the dynamic standard model. The dynamic standard model explicitly accounts for changes in the standard of accountability over time. The difference between the two series increases as a function of time.

F The Changing Standard of Accountability: Additional Plots and Tables

Figure 5 captures the increasing disagreement between the latent variables estimates generated from the Dynamic Standard Model and those from the Constant Standard Model (1976-2010). The standard of accountability affects each of the standards-based variables differently. These differential effects are captured in Figure 6, 7, 8, 9, 10, 11, 12, and 13. Each of these figures display four or more panels that illustrate how the changing standard of accountability affects the standards-based repression variables. The upper most left panel displays temporal change in the item difficulty cut-points from the dynamic standard model. Notice for example that there is substantial change for the CIRI torture variable but very little for the CIRI political imprisonment variable. To get an overall view of the effect the standard of accountability has on changes to the coding the standards-based data, consider Table 2 and Table 3. Each row in the table reports slope coefficients and R^2 statistics from binary linear regression models (one model per row). Each set of item difficulty cut-points $\alpha_{t,k}$ are regressed on an index of time periods t, where $t = 1, \ldots, T$. Recall that higher values on the CIRI variables indicate greater respect (less repression). Higher values on the other variables indicate less respect (greater repression). The positive signed coefficients on several of the CIRI variables indicate that as t increases the difficulty cut-points also increase just as the figures for the CIRI torture variable suggest. An increase in the difficulty cut-points translates directly into a change in the probability of being classified as a 0, 1, or 2 on the original CIRI variables such that begin classified as 0 (e.g., frequent torture) becomes more likely and 2 (e.g. no torture) becomes less likely over time. The lack of results for these two variables is actually quite encouraging for the plausibility of the dynamic standard model. In effect, these two variables in addition to the five event-based indicators acted as a baseline for the model so that both the overall level of repression and the changing standard of accountability could be estimated simultaneously. These results help to alleviate concern that the changing standard of accountability is an unwanted artifact rather than a theoretically

specified feature of the model.

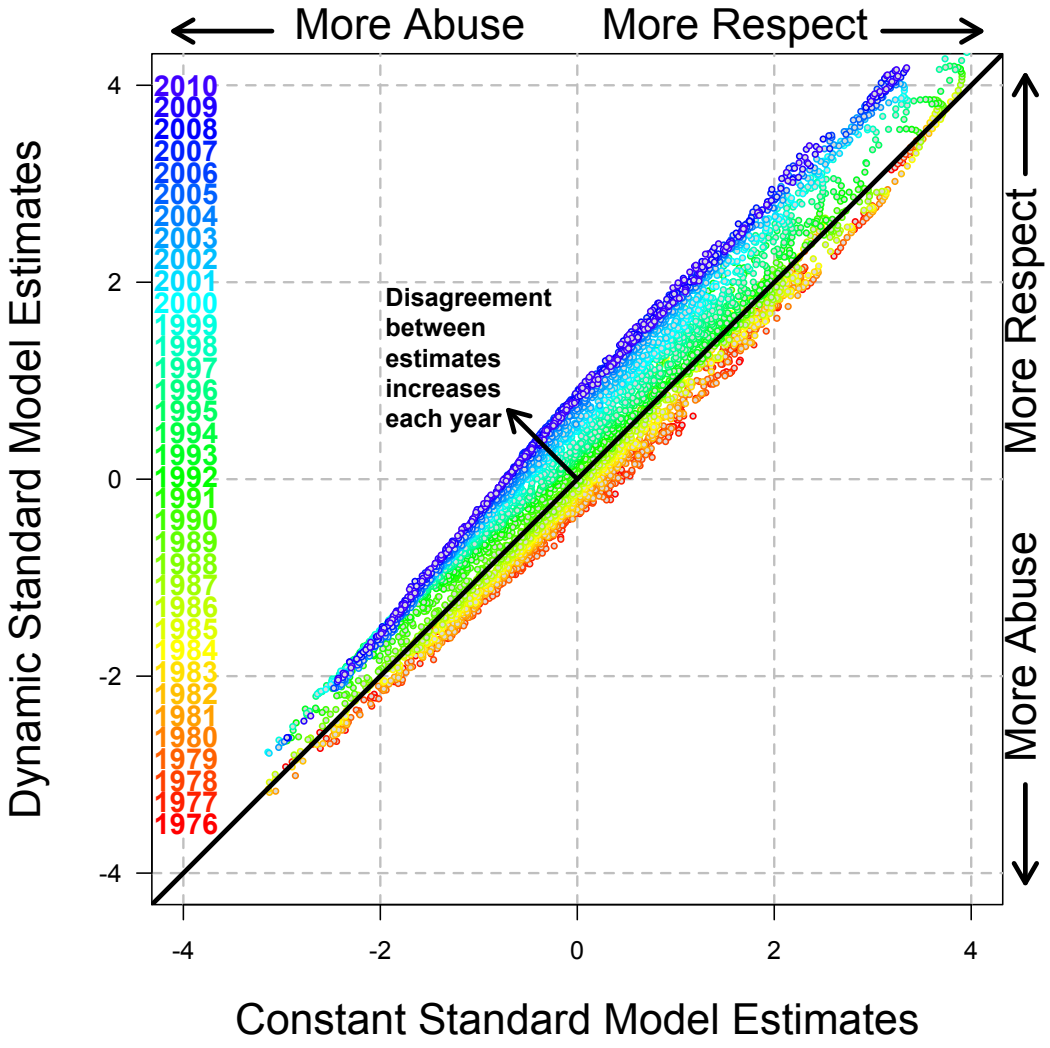

Figure 5: Relationship between the latent variable estimates generated from the Dynamic Standard Model on the y-axis and the estimates generated from the Constant Standard Model on the x-axis (1976-2010). The 45-degree line represents perfect agreement between the two estimates. Disagreement between the two sets of estimates increases as a function of time.

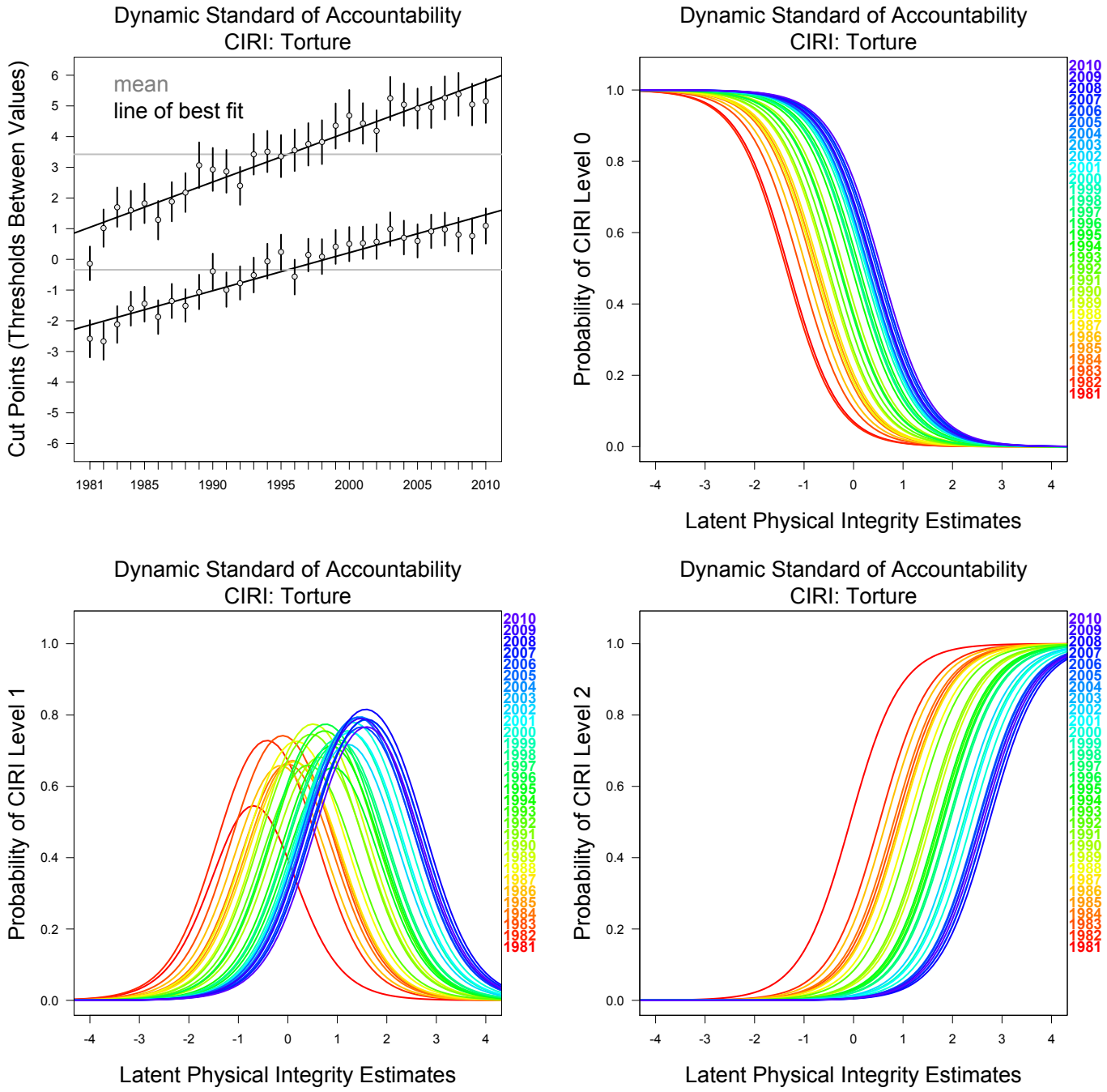

Figure 6: An increase in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 0, 1, or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent torture) becomes more likely and 2 (e.g. no torture) becomes less likely as a function of time. See Section G for the posterior estimates of these parameters.

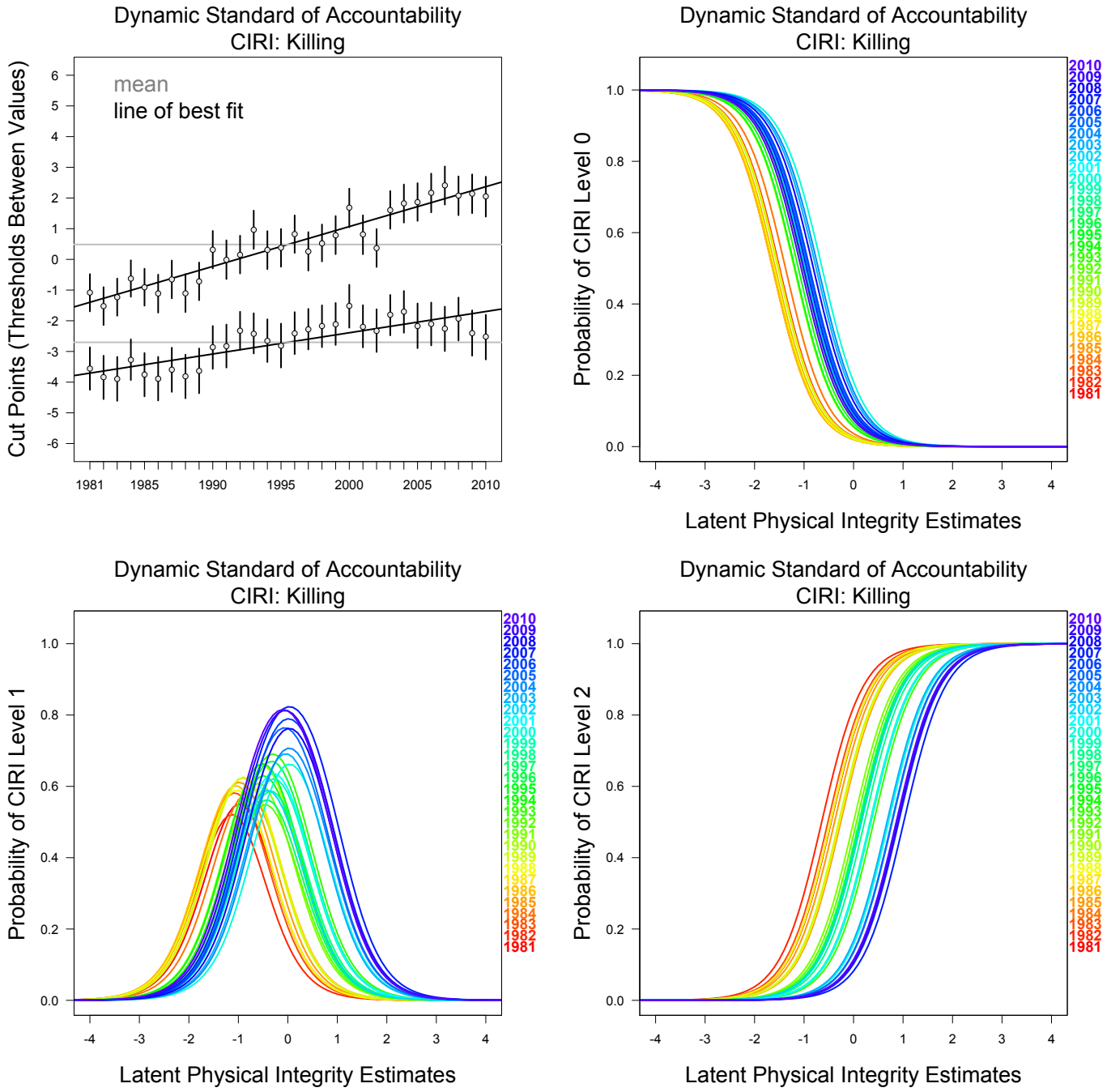

Figure 7: An increase in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 0, 1 or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent extrajudicial killing) becomes more likely and 2 (e.g. no extrajudicial killing) becomes less likely as a function of time. See Section G for the posterior estimates of these parameters.

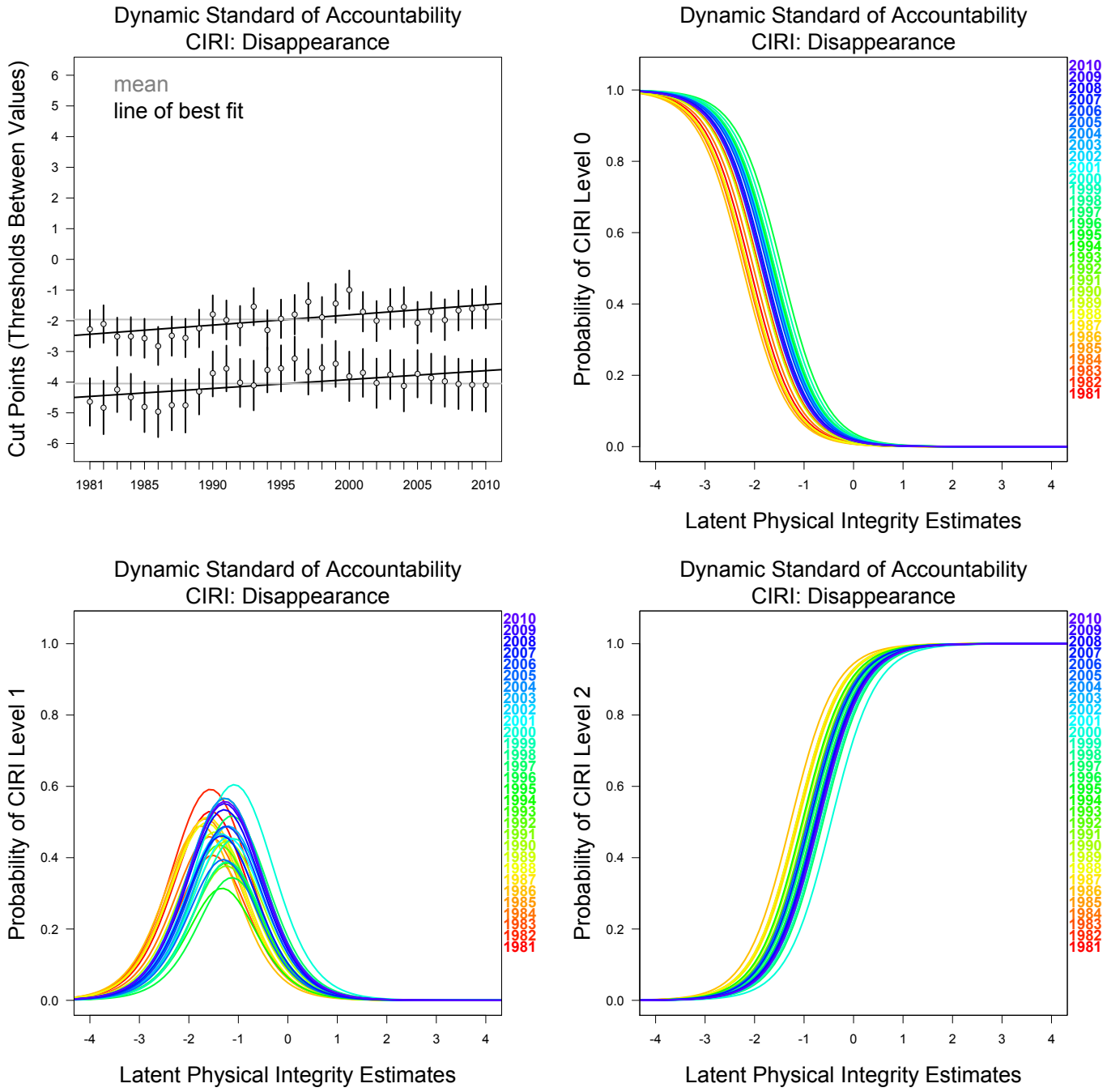

Figure 8: An increase in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 0, 1 or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent disappearances) becomes more likely and 2 (e.g. no disapperances) becomes less likely as a function of time. See Section G for the posterior estimates of these parameters.

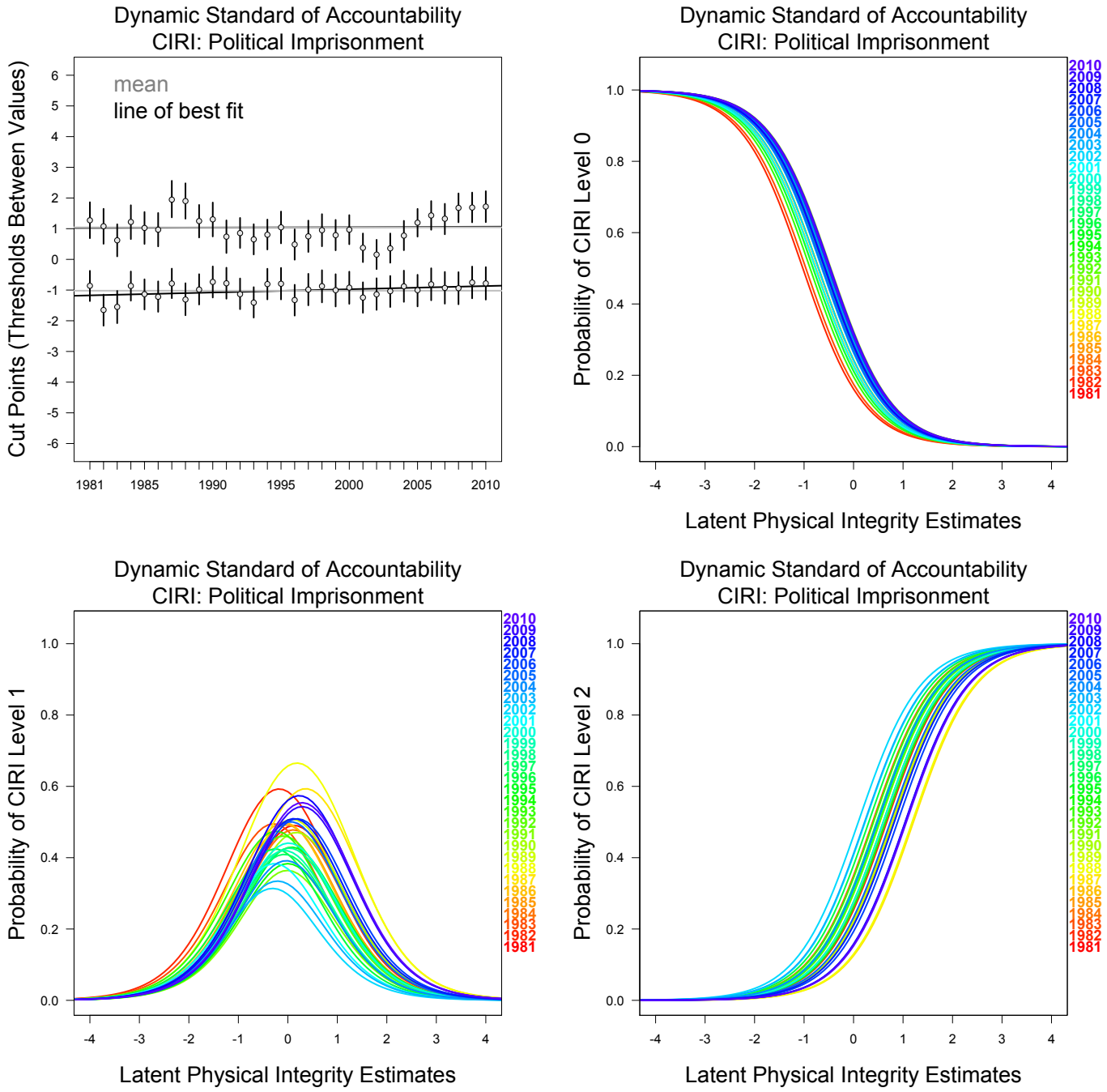

Figure 9: Very little change occurs over time for the difficulty cut-points in the upper left panel. Therefore, the probability of being classified as a 0, 1, or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent political imprisonment) or a 2 (e.g. no political imprisonment) does not vary as a function of time. See Section G for the posterior estimates of these parameters.

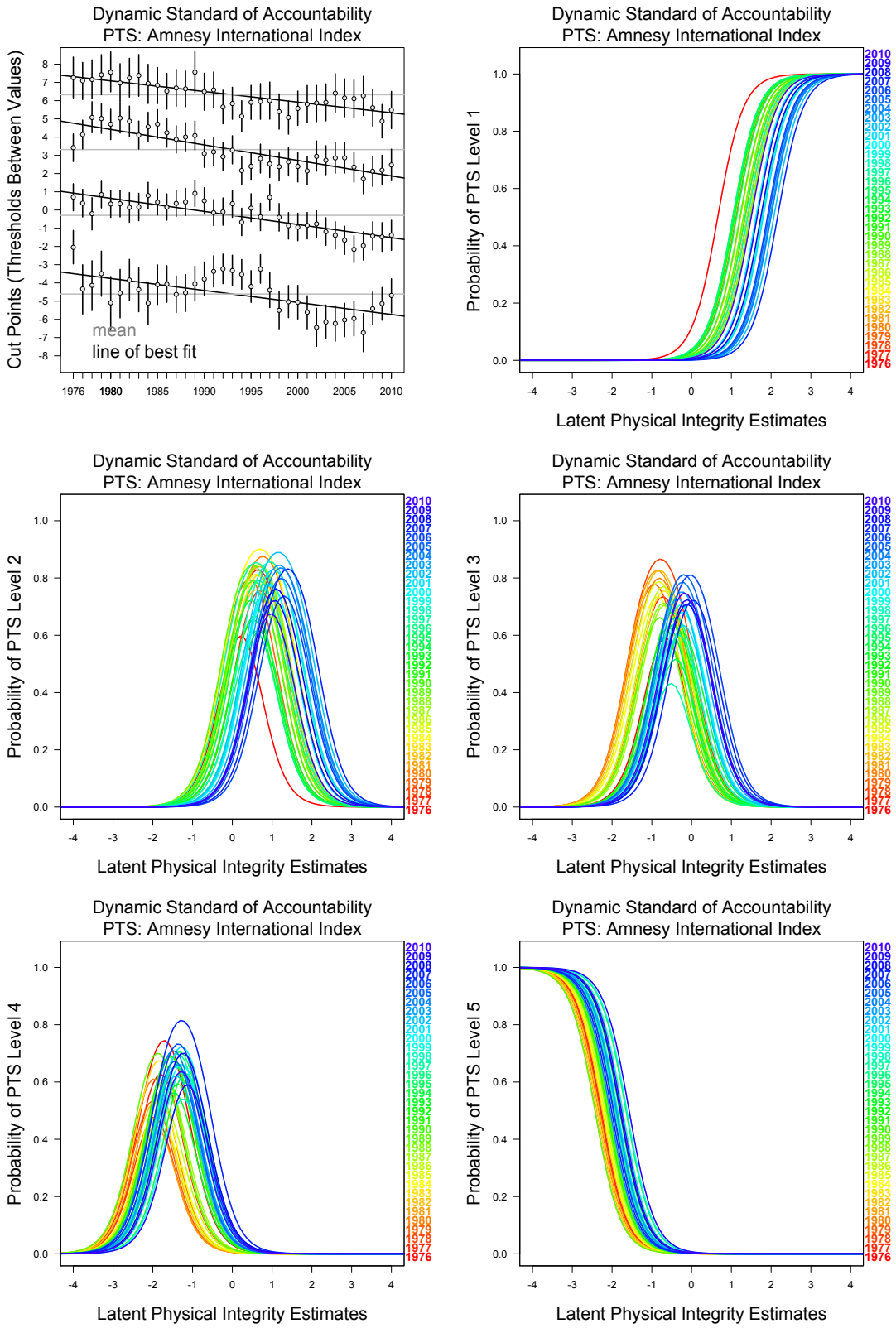

Figure 10: A decrease in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original PTS Amnesty Index such that begin classified as 5 (e.g., frequent abuse) becomes more likely and 1 (e.g. no abuse) becomes less likely as a function of time. See Section G for the posterior estimates of these parameters.

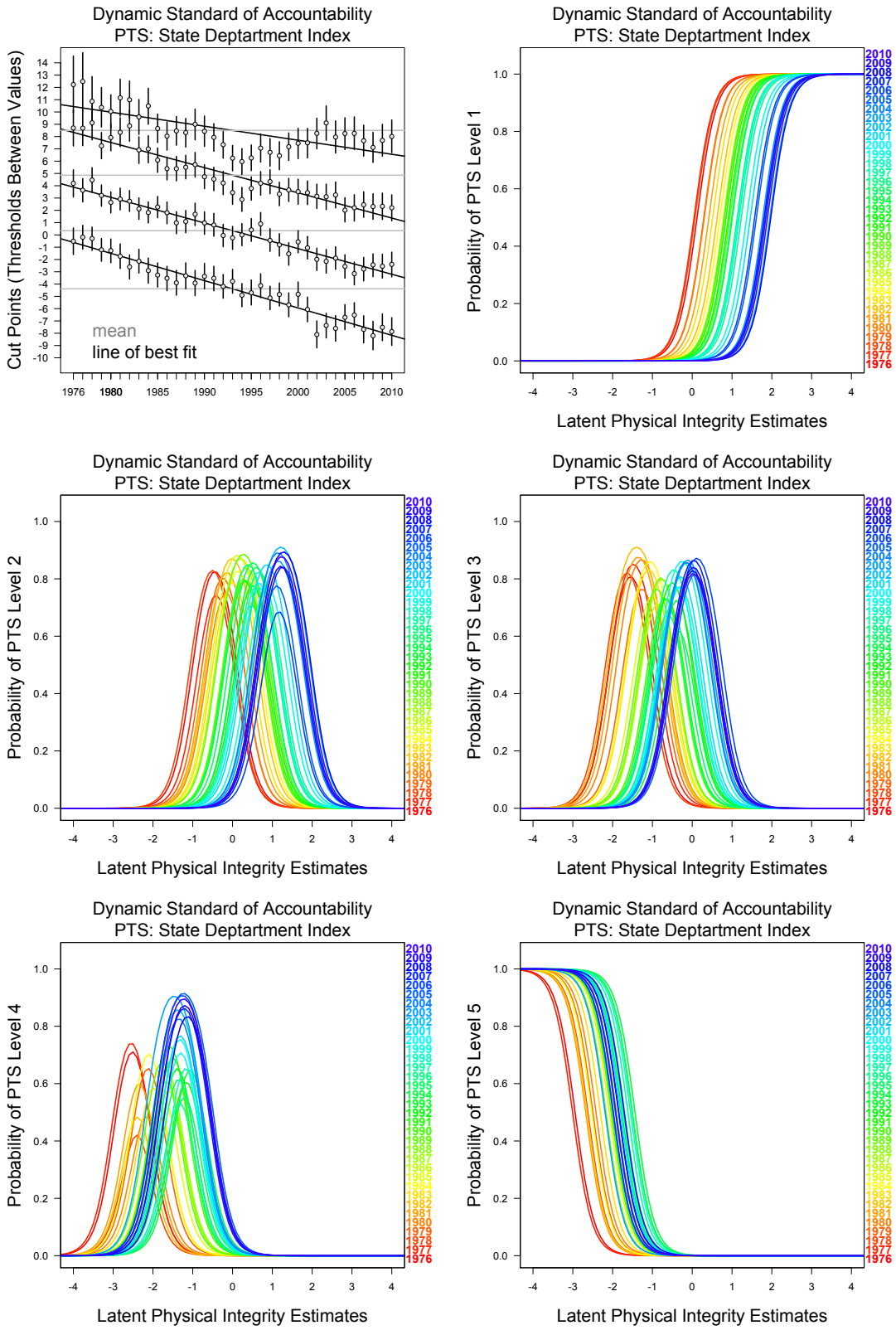

Figure 11: A decrease in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original PTS State Department Index such that begin classified as 5 (e.g., frequent abuse) becomes more likely and 1 (e.g. no abuse) becomes less likely as a function of time. See Section G for the posterior estimates of these parameters.

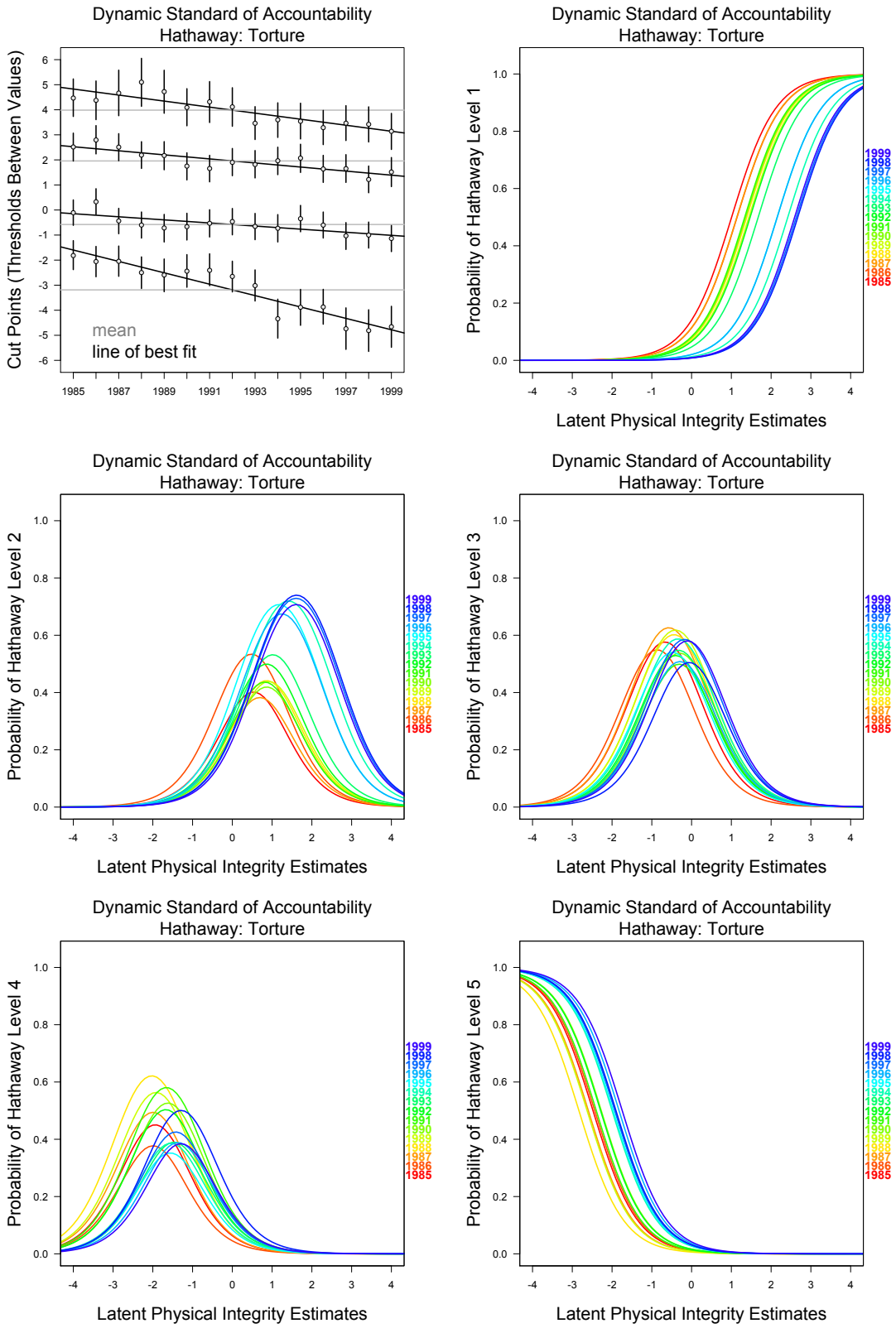

Figure 12: A decrease in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original Hathaway Torture Index such that begin classified as 5 (e.g., frequent tortyre) becomes more likely and 1 (e.g. no torture) becomes less likely as a function of time. See Section G for the posterior estimates of these parameters.

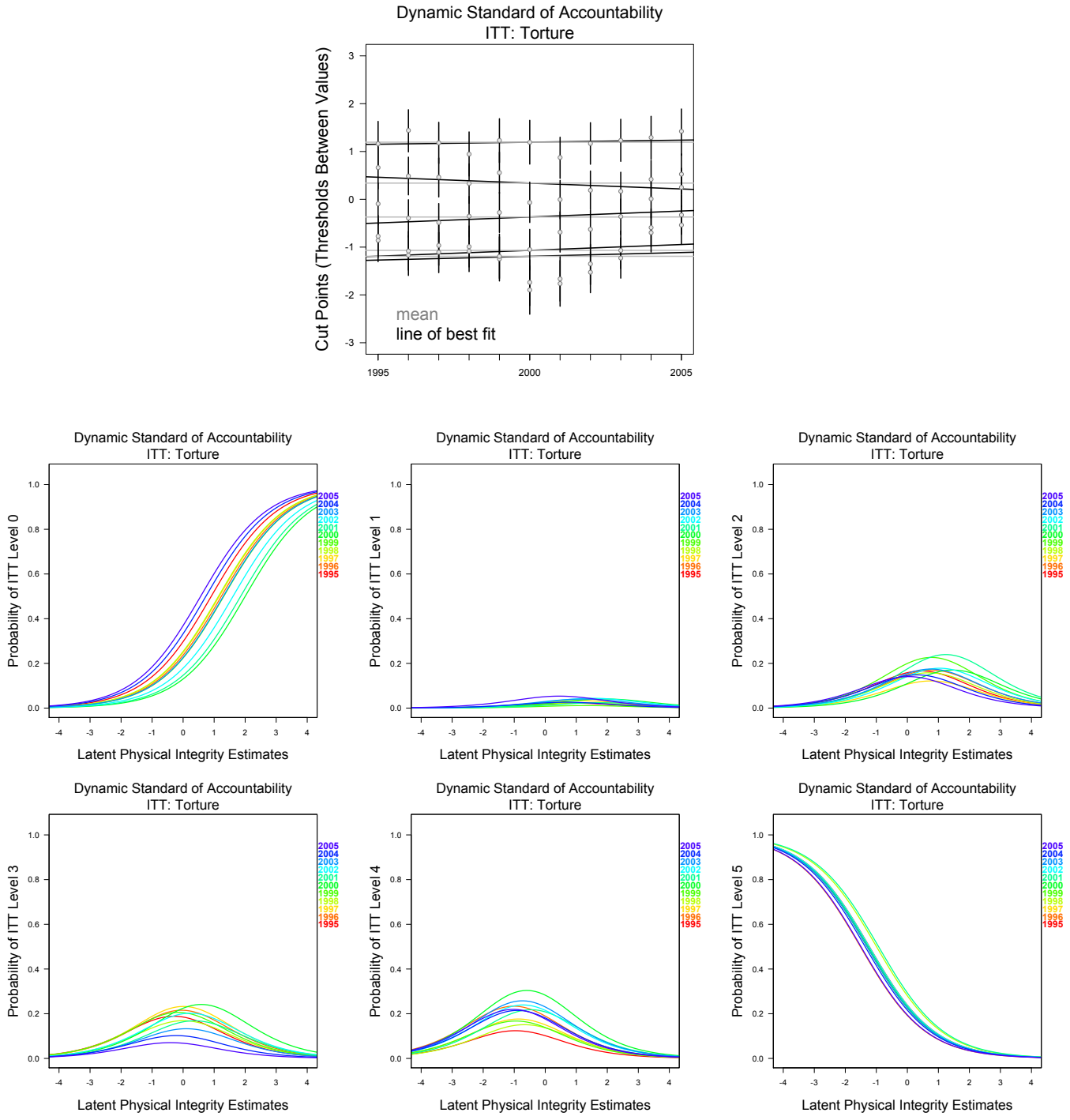

Figure 13: Very little change occurs over time for the difficulty cut-points in the upper. Therefore, the probability of being classified as a 0, 1, 2, 3, 4, or 5 on the original ITT Torture Index does not vary over time. See Section G for the posterior estimates of these parameters.

Item Difficulty Cut-Points		Coefficient [95%CI]	R^2 [95%CI]
CIRI: torture			
Threshold Between 0 and 1	$\alpha_{t,1}$	0.1239 [0.1123, 0.1357]	0.8626 [0.7998, 0.9135]
Threshold Between 1 and 2	$\alpha_{t,2}$	0.1641 [0.1501, 0.1782]	0.8809 [0.8297, 0.9257]
CIRI: killing			
Threshold Between 0 and 1	$\alpha_{t,1}$	0.0695 [0.0548, 0.0841]	0.5711 [0.4142, 0.7125]
Threshold Between 1 and 2	$\alpha_{t,2}$	0.1299 [0.1168, 0.1433]	0.8459 [0.7760, 0.9005]
CIRI: imprisonment **no significant trend**			
Threshold Between 0 and 1	$\alpha_{t,1}$	0.0107 [-0.0003, 0.0213]	0.0726 [0.0011, 0.2557]
Threshold Between 1 and 2	$\alpha_{t,2}$	0.0013 [-0.0093, 0.0121]	0.0040 [0.0000, 0.0478]
CIRI: disappearance			
Threshold Between 0 and 1	$\alpha_{t,1}$	0.0291 [0.0122, 0.0462]	0.1748 [0.0346, 0.3802]
Threshold Between 1 and 2	$\alpha_{t,2}$	0.0331 [0.0196, 0.0462]	0.3013 [0.1257, 0.5011]

Table 2: Binary regression slope coefficients and R^2 statistics (one model per row). Each set of item difficulty cut-points $\alpha_{t,k}$ or thresholds between values are regressed on the index t, where $t = 1, \ldots, T$ and indexes time periods. The number of difficulty cut-points per item is $K_j - 1$, where K_j is the number of ordinal values for that variable j. Recall that higher values on the CIRI variable indicate greater respect (less repression), whereas higher values on the other variables indicate less respect (greater repression). The positive signed coefficients on cut-points indicate that as the time period index increases the difficulty cut-points also increase. An increase in the difficulty cut-points translates directly into a change in the probability of being classified as a 0, 1, or 2 on the original CIRI variables such that begin classified as 0 (e.g., frequent abuse) becomes more likely and 2 (e.g. no abuse) becomes less likely as a function of time. Credible intervals are calculated by running 1000 binary regressions taking a new draw from the posterior of the difficulty cut-point every iteration and saving the estimate of the slope coefficient and R^2 statistic. See Section G for parameter estimates of each of the item difficulty cut-points $\alpha_{t,k}$.

Item Difficulty Cut-Points		Coefficient [95%CI]	R^2 [95%CI]
PTS: State			
Threshold Between 1 and 2	$\alpha_{t,1}$	-0.2224 [-0.2392, -0.2054]	0.8949 [0.8516, 0.9314]
Threshold Between 2 and 3	$\alpha_{t,2}$	-0.2087 [-0.2253, -0.1931]	0.9132 [0.8744, 0.9449]
Threshold Between 3 and 4	$\alpha_{t,3}$	-0.2050 [-0.2247, -0.1854]	0.8443 [0.7852, 0.8928]
Threshold Between 4 and 5	$\alpha_{t,4}$	-0.1141 [-0.1417, -0.0879]	0.4180 [0.2955, 0.5354]
PTS: Amnesty			
Threshold Between 1 and 2	$\alpha_{t,1}$	-0.0656 [-0.0840, -0.0481]	0.3242 [0.1883, 0.4773]
Threshold Between 2 and 3	$\alpha_{t,2}$	-0.0721 [-0.0848, -0.0588]	0.6034 [0.4693, 0.7236]
Threshold Between 3 and 4	$\alpha_{t,3}$	-0.0850 [-0.0988, -0.0715]	0.6402 [0.5233, 0.7484]
Threshold Between 4 and 5	$\alpha_{t,4}$	-0.0580 [-0.0758, -0.0406]	0.4310 [0.2469, 0.6070]
Hathaway: torture			
Threshold Between 1 and 2	$\alpha_{t,1}$	-0.2278 [-0.2717, -0.1837]	0.8154 [0.6853, 0.9089]
Threshold Between 2 and 3	$\alpha_{t,2}$	-0.0618 [-0.0933, -0.0303]	0.3878 [0.1172, 0.6716]
Threshold Between 3 and 4	$\alpha_{t,3}$	-0.0809 [-0.1133, -0.0489]	0.5280 [0.2504, 0.7691]
Threshold Between 4 and 5	$\alpha_{t,4}$	-0.1198 [-0.1651, -0.0765]	0.5701 [0.3039, 0.7976]
ITT: torture **no significant trend**			
Threshold Between 0 and 1	$\alpha_{t,1}$	0.0153 [-0.0245, 0.0564]	0.0166 [0.0000, 0.1589]
Threshold Between 1 and 2	$\alpha_{t,2}$	0.0249 [-0.0147, 0.0649]	0.0324 [0.0001, 0.2019]
Threshold Between 2 and 3	$\alpha_{t,3}$	0.0248 [-0.0141, 0.0633]	0.0433 [0.0001, 0.2538]
Threshold Between 3 and 4	$\alpha_{t,4}$	-0.0239 [-0.0647, 0.0152]	0.0709 [0.0003, 0.4115]
Threshold Between 4 and 5	$\alpha_{t,5}$	0.0084 [-0.0361, 0.0530]	0.0397 [0.0001, 0.3626]

Table 3: Binary regression slope coefficients and R^2 statistics (one model per row). Each set of item difficulty cut-points $\alpha_{t,k}$ or thresholds between values are regressed on the index t, where $t = 1, \ldots, T$ and indexes time periods. The number of difficulty cut-points per item is $K_j - 1$, where K_j is the number of ordinal values for that variable j. These repression variables are coded in reverse with respect to CIRI, therefore a decrease in the difficulty cut-points for these variables translates into similar changes in the probability of being classified relative to the CIRI variables. Credible intervals are calculated by running 1000 binary regressions taking a new draw from the posterior of the difficulty cut-point every iteration and saving the estimate of the slope coefficient and R^2 statistic. See Section G for parameter estimates of each of the item difficulty cut-points $\alpha_{t,k}$ and for individual probabilities in the form $Pr(\alpha_{t-1,k} < \alpha_{t,k})$.

G Dynamic Standard Model Cut-points for the Standards-Based Response Variables

Model parameters are displayed for the posterior estimates of the dynamic cut-points estimated from the dynamic standard model. These parameters are regressed on the index of time. Results for these regression are displayed in Table 2 and Table 3 above.

t	$\alpha_{t,1}$	$Pr(\alpha_{t-1,1} < \alpha_{t,1})$	$\alpha_{t,2}$	$Pr(\alpha_{t-1,2} < \alpha_{t,2})$
1981	-4.641 (0.404)		-2.281 (0.309)	
1982	-4.837 (0.437)	0.366	-2.116 (0.314)	0.642
1983	-4.237 (0.381)	0.869	-2.511 (0.316)	0.180
1984	-4.492 (0.393)	0.309	-2.508 (0.318)	0.500
1985	-4.819 (0.424)	0.328	-2.571 (0.325)	0.410
1986	-4.964 (0.423)	0.396	-2.825 (0.324)	0.308
1987	-4.754 (0.422)	0.595	-2.484 (0.327)	0.772
1988	-4.75 (0.435)	0.544	-2.564 (0.321)	0.429
1989	-4.304 (0.386)	0.778	-2.253 (0.316)	0.757
1990	-3.71 (0.378)	0.877	-1.79 (0.315)	0.849
1991	-3.558 (0.385)	0.634	-1.972 (0.325)	0.333
1992	-4.011 (0.404)	0.181	-2.142 (0.327)	0.353
1993	-4.109 (0.403)	0.436	-1.54 (0.309)	0.914
1994	-3.607 (0.373)	0.817	-2.306 (0.331)	0.050
1995	-3.555 (0.385)	0.514	-1.929 (0.320)	0.809
1996	-3.229 (0.359)	0.745	-1.794 (0.327)	0.610
1997	-3.661 (0.382)	0.186	-1.371 (0.317)	0.843
1998	-3.535 (0.384)	0.581	-1.892 (0.329)	0.115
1999	-3.398 (0.393)	0.658	-1.442 (0.317)	0.832
2000	-3.814 (0.410)	0.223	-1.011 (0.321)	0.854
2001	-3.684 (0.400)	0.584	-1.709 (0.328)	0.057
2002	-4.018 (0.417)	0.290	-1.993 (0.333)	0.270
2003	-3.749 (0.414)	0.690	-1.612 (0.325)	0.797
2004	-4.117 (0.435)	0.264	-1.545 (0.334)	0.566
2005	-3.732 (0.398)	0.704	-2.063 (0.352)	0.126
2006	-3.853 (0.423)	0.437	-1.72 (0.330)	0.763
2007	-3.975 (0.423)	0.462	-1.978 (0.337)	0.282
2008	-4.054 (0.423)	0.511	-1.668 (0.338)	0.732
2009	-4.09 (0.421)	0.451	-1.607 (0.329)	0.568
2010	-4.087 (0.443)	0.453	-1.568 (0.347)	0.553

Table 4: Item difficulty posterior estimates and standard deviations for the CIRI disappearance item. These parameters are displayed visually in the main document. Table 3 displays information about the linear trend of the change in the cut point values in addition to the probability that one cut point is different than the one preceding it displayed here as: $Pr(\alpha_{t-1,k} < \alpha_{t,k})$.

t	$\alpha_{t,1}$	$Pr(\alpha_{t-1,1} < \alpha_{t,1})$	$\alpha_{t,2}$	$Pr(\alpha_{t-1,2} < \alpha_{t,2})$
1981	(3.557) -0.36		-1.08 (0.311)	
1982	(3.844) -0.365	0.295	-1.532 (0.311)	0.131
1983	(3.894) -0.365	0.467	-1.233 (0.310)	0.741
1984	(3.272) -0.337	0.895	-0.62 (0.303)	0.919
1985	(3.762) -0.361	0.171	-0.916 (0.312)	0.233
1986	(3.896) -0.364	0.402	-1.109 (0.313)	0.324
1987	(3.593) -0.363	0.676	-0.655 (0.313)	0.879
1988	(3.812) -0.357	0.342	-1.105 (0.313)	0.149
1989	(3.641) -0.372	0.698	-0.717 (0.315)	0.775
1990	(2.855) -0.348	0.946	0.315 (0.319)	0.987
1991	(2.824) -0.354	0.493	0 (0.327)	0.285
1992	(2.326) -0.324	0.849	0.143 (0.310)	0.632
1993	(2.430) -0.333	0.431	0.968 (0.318)	0.965
1994	(2.654) -0.351	0.308	0.31 (0.317)	0.084
1995	(2.818) -0.364	0.347	0.378 (0.311)	0.562
1996	(2.412) -0.342	0.804	0.834 (0.315)	0.857
1997	(2.284) -0.34	0.640	0.261 (0.319)	0.098
1998	(2.165) -0.345	0.587	0.526 (0.312)	0.729
1999	(2.120) -0.347	0.550	0.788 (0.316)	0.728
2000	(1.511) -0.352	0.900	1.677 (0.322)	0.977
2001	(2.197) -0.351	0.070	0.812 (0.317)	0.032
2002	(2.332) -0.353	0.401	0.372 (0.318)	0.170
2003	(1.798) -0.344	0.859	1.607 (0.315)	0.998
2004	(1.696) -0.348	0.539	1.832 (0.309)	0.711
2005	(2.169) -0.37	0.174	1.859 (0.310)	0.519
2006	(2.110) -0.364	0.552	2.172 (0.324)	0.737
2007	(2.256) -0.366	0.413	2.415 (0.324)	0.706
2008	(1.932) -0.352	0.752	2.079 (0.321)	0.223
2009	(2.404) -0.374	0.172	2.141 (0.328)	0.548
2010	(2.517) -0.383	0.412	2.054 (0.336)	0.436

Table 5: Item difficulty posterior estimates and standard deviations for the CIRI extrajudicial killing item. These parameters are displayed visually in the main document. Table 3 displays information about the linear trend of the change in the cut point values in addition to the probability that one cut point is different than the one preceding it displayed here as: $Pr(\alpha_{t-1,k} < \alpha_{t,k})$.

t	$\alpha_{t,1}$	$Pr(\alpha_{t-1,1} < \alpha_{t,1})$	$\alpha_{t,2}$	$Pr(\alpha_{t-1,2} < \alpha_{t,2})$
1981	-2.585 (0.308)		-0.136 (0.285)	
1982	-2.675 (0.306)	0.416	1.029 (0.318)	0.996
1983	-2.118 (0.296)	0.896	1.704 (0.327)	0.932
1984	-1.598 (0.290)	0.896	1.591 (0.331)	0.403
1985	-1.439 (0.282)	0.665	1.822 (0.329)	0.686
1986	-1.87 (0.285)	0.142	1.289 (0.312)	0.113
1987	-1.353 (0.282)	0.893	1.885 (0.326)	0.910
1988	-1.51 (0.274)	0.361	2.176 (0.329)	0.732
1989	-1.066 (0.292)	0.860	3.062 (0.374)	0.962
1990	-0.39 (0.291)	0.954	2.941 (0.360)	0.401
1991	-0.993 (0.290)	0.066	2.863 (0.355)	0.439
1992	-0.775 (0.283)	0.713	2.397 (0.319)	0.160
1993	-0.52 (0.288)	0.740	3.424 (0.340)	0.988
1994	-0.06 (0.287)	0.879	3.513 (0.348)	0.601
1995	0.236 (0.290)	0.790	3.356 (0.347)	0.361
1996	-0.567 (0.289)	0.020	3.559 (0.348)	0.682
1997	0.137 (0.280)	0.963	3.757 (0.353)	0.662
1998	0.088 (0.286)	0.476	3.82 (0.360)	0.559
1999	0.413 (0.281)	0.785	4.359 (0.374)	0.839
2000	0.504 (0.277)	0.614	4.679 (0.425)	0.718
2001	0.525 (0.275)	0.529	4.436 (0.332)	0.318
2002	0.573 (0.284)	0.552	4.186 (0.351)	0.306
2003	0.994 (0.275)	0.861	5.247 (0.360)	0.984
2004	0.701 (0.279)	0.228	5.04 (0.354)	0.350
2005	0.603 (0.274)	0.397	4.928 (0.342)	0.420
2006	0.907 (0.277)	0.793	4.955 (0.338)	0.553
2007	0.99 (0.279)	0.584	5.258 (0.356)	0.724
2008	0.802 (0.283)	0.306	5.381 (0.359)	0.611
2009	0.761 (0.294)	0.456	5.055 (0.352)	0.241
2010	1.097 (0.295)	0.798	5.147 (0.371)	0.589

Table 6: Item difficulty posterior estimates and standard deviations for the CIRI torture item. These parameters are displayed visually in the main document. Table 3 displays information about the linear trend of the change in the cut point values in addition to the probability that one cut point is different than the one preceding it displayed here as: $Pr(\alpha_{t-1,k} < \alpha_{t,k})$.

t	$\alpha_{t,1}$	$Pr(\alpha_{t-1,1} < \alpha_{t,1})$	$\alpha_{t,2}$	$Pr(\alpha_{t-1,2} < \alpha_{t,2})$
1981	-0.867 (0.253)		1.28 (0.295)	
1982	-1.648 (0.256)	0.016	1.08 (0.290)	0.311
1983	-1.55 (0.264)	0.608	0.62 (0.268)	0.116
1984	-0.864 (0.252)	0.968	1.218 (0.277)	0.936
1985	-1.133 (0.256)	0.228	1.022 (0.271)	0.309
1986	-1.22 (0.253)	0.392	0.969 (0.273)	0.429
1987	-0.788 (0.247)	0.893	1.944 (0.306)	0.994
1988	-1.306 (0.263)	0.079	1.904 (0.291)	0.453
1989	-0.985 (0.253)	0.802	1.253 (0.271)	0.051
1990	-0.738 (0.257)	0.762	1.306 (0.283)	0.560
1991	-0.78 (0.259)	0.465	0.747 (0.272)	0.071
1992	-1.138 (0.253)	0.150	0.855 (0.252)	0.620
1993	-1.412 (0.254)	0.211	0.649 (0.252)	0.314
1994	-0.811 (0.254)	0.954	0.805 (0.253)	0.681
1995	-0.796 (0.259)	0.532	1.042 (0.262)	0.754
1996	-1.317 (0.256)	0.075	0.495 (0.253)	0.064
1997	-0.986 (0.261)	0.828	0.753 (0.260)	0.775
1998	-0.865 (0.264)	0.643	0.96 (0.264)	0.702
1999	-1.003 (0.264)	0.363	0.788 (0.253)	0.310
2000	-0.922 (0.267)	0.591	0.972 (0.251)	0.706
2001	-1.239 (0.255)	0.190	0.377 (0.244)	0.041
2002	-1.145 (0.265)	0.632	0.154 (0.245)	0.281
2003	-1.032 (0.260)	0.605	0.36 (0.245)	0.740
2004	-0.879 (0.259)	0.649	0.776 (0.247)	0.904
2005	-1.008 (0.262)	0.345	1.193 (0.236)	0.886
2006	-0.815 (0.268)	0.693	1.431 (0.242)	0.777
2007	-0.924 (0.263)	0.399	1.324 (0.250)	0.373
2008	-0.937 (0.269)	0.514	1.676 (0.242)	0.833
2009	-0.751 (0.266)	0.713	1.684 (0.245)	0.481
2010	-0.782 (0.272)	0.462	1.716 (0.259)	0.552

Table 7: Item difficulty posterior estimates and standard deviations for the CIRI political imprisonment item. These parameters are displayed visually in the main document. Table 3 displays information about the linear trend of the change in the cut point values in addition to the probability that one cut point is different than the one preceding it displayed here as: $Pr(\alpha_{t-1,k} < \alpha_{t,k})$. Though there are some changes in terms of the $Pr(\alpha_{t-1,k} < \alpha_{t,k})$ the change is not systematic as it is with the other three CIRI variables as displayed in Table 3.

t	$\alpha_{t,1}$	$\alpha_{t,2}$	$\alpha_{t,3}$	$\alpha_{t,4}$
1976	-2.052 (0.478)	0.702 (0.384)	3.420 (0.392)	7.264 (0.582)
1977	-4.357 (0.687)	0.380 (0.395)	4.140 (0.407)	7.091 (0.524)
1978	-4.140 (0.684)	-0.192 (0.440)	5.075 (0.431)	7.157 (0.534)
1979	-3.486 (0.621)	0.849 (0.376)	5.019 (0.418)	7.408 (0.580)
1980	-5.093 (0.722)	0.325 (0.385)	4.713 (0.415)	7.556 (0.563)
1981	-4.542 (0.666)	0.346 (0.376)	5.054 (0.432)	6.980 (0.521)
1982	-3.827 (0.576)	0.156 (0.411)	4.874 (0.421)	7.244 (0.546)
1983	-4.360 (0.611)	0.169 (0.388)	4.109 (0.395)	7.391 (0.591)
1984	-5.115 (0.601)	0.794 (0.371)	4.551 (0.405)	6.935 (0.544)
1985	-4.108 (0.543)	0.439 (0.367)	4.697 (0.405)	6.834 (0.500)
1986	-4.066 (0.516)	0.156 (0.370)	4.243 (0.402)	6.516 (0.504)
1987	-4.626 (0.501)	0.365 (0.371)	3.875 (0.393)	6.705 (0.511)
1988	-4.545 (0.518)	0.439 (0.372)	4.006 (0.403)	6.644 (0.509)
1989	-4.057 (0.496)	0.909 (0.370)	4.094 (0.407)	7.572 (0.573)
1990	-3.792 (0.465)	0.506 (0.365)	3.096 (0.387)	6.480 (0.501)
1991	-3.373 (0.439)	-0.146 (0.374)	3.186 (0.395)	6.590 (0.506)
1992	-3.227 (0.414)	-0.097 (0.382)	2.917 (0.385)	5.650 (0.465)
1993	-3.315 (0.423)	0.340 (0.378)	3.286 (0.397)	5.833 (0.456)
1994	-3.532 (0.411)	-0.668 (0.381)	2.167 (0.390)	5.137 (0.437)
1995	-4.204 (0.443)	0.101 (0.383)	2.390 (0.389)	5.909 (0.481)
1996	-3.235 (0.405)	-0.369 (0.383)	2.823 (0.398)	5.957 (0.495)
1997	-4.400 (0.492)	0.688 (0.374)	2.537 (0.383)	5.996 (0.528)
1998	-5.522 (0.520)	-0.398 (0.379)	2.381 (0.395)	5.398 (0.490)
1999	-5.040 (0.452)	-0.866 (0.369)	2.646 (0.393)	5.073 (0.469)
2000	-5.060 (0.467)	-0.936 (0.380)	2.388 (0.413)	5.574 (0.488)
2001	-5.610 (0.463)	-0.859 (0.371)	2.138 (0.396)	5.800 (0.522)
2002	-6.443 (0.527)	-0.758 (0.371)	2.942 (0.427)	5.863 (0.505)
2003	-6.146 (0.483)	-1.190 (0.389)	2.723 (0.412)	5.897 (0.521)
2004	-6.221 (0.494)	-1.397 (0.381)	2.842 (0.428)	6.390 (0.537)
2005	-6.033 (0.468)	-1.651 (0.389)	2.874 (0.427)	6.138 (0.537)
2006	-5.938 (0.481)	-2.161 (0.389)	2.359 (0.419)	6.104 (0.537)
2007	-6.733 (0.527)	-1.958 (0.394)	1.698 (0.420)	6.269 (0.550)
2008	-5.423 (0.455)	-1.420 (0.396)	2.122 (0.420)	5.601 (0.518)
2009	-5.134 (0.456)	-1.488 (0.401)	2.178 (0.426)	4.894 (0.476)
2010	-4.675 (0.473)	-1.390 (0.418)	2.479 (0.440)	5.495 (0.518)

Table 8: Item difficulty posterior estimates and standard deviations for the PTS Amnesty item. These parameters are displayed visually in the main document.

t	$Pr(\alpha_{t-1,1} > \alpha_{t,1})$	$Pr(\alpha_{t-1,2} > \alpha_{t,2})$	$Pr(\alpha_{t-1,3} > \alpha_{t,3})$	$Pr(\alpha_{t-1,4} > \alpha_{t,4})$
1976				
1977	0.994	0.746	0.107	0.530
1978	0.391	0.841	0.046	0.561
1979	0.261	0.034	0.497	0.440
1980	0.909	0.833	0.699	0.468
1981	0.351	0.440	0.318	0.714
1982	0.362	0.536	0.645	0.341
1983	0.857	0.409	0.896	0.453
1984	0.696	0.162	0.214	0.750
1985	0.064	0.776	0.422	0.537
1986	0.456	0.680	0.759	0.687
1987	0.774	0.358	0.733	0.411
1988	0.445	0.446	0.410	0.510
1989	0.171	0.240	0.457	0.121
1990	0.373	0.766	0.961	0.919
1991	0.310	0.862	0.419	0.435
1992	0.338	0.499	0.697	0.913
1993	0.572	0.209	0.264	0.391
1994	0.660	0.972	0.981	0.864
1995	0.865	0.075	0.368	0.129
1996	0.108	0.756	0.210	0.460
1997	0.940	0.052	0.656	0.475
1998	0.910	0.985	0.654	0.797
1999	0.354	0.737	0.308	0.669
2000	0.430	0.616	0.692	0.240
2001	0.763	0.512	0.720	0.333
2002	0.882	0.406	0.127	0.390
2003	0.340	0.787	0.618	0.578
2004	0.596	0.627	0.345	0.351
2005	0.421	0.719	0.440	0.632
2006	0.427	0.829	0.776	0.525
2007	0.848	0.377	0.872	0.419
2008	0.026	0.170	0.279	0.800
2009	0.392	0.500	0.472	0.839
2010	0.253	0.410	0.313	0.192

Table 9: The probability that one PTS Amnesty cut point is different than the one preceding it in time is displayed here as: $Pr(\alpha_{t-1,k} > \alpha_{t,k})$. These inequality tests are reversed relative to CIRI because the ordering of the original scales are reversed themselves so the PTS cut points should decrease over time whereas the CIRI cut-points should increase. Note that there are many yearly differences, which are highly probably, which corroborates the non-zero linear trend for this as displayed in Table 3 above.

t	$\alpha_{t,1}$	$\alpha_{t,2}$	$\alpha_{t,3}$	$\alpha_{t,4}$
1976	-0.505 (0.560)	4.192 (0.561)	8.678 (0.734)	12.232 (1.181)
1977	-0.194 (0.495)	3.634 (0.488)	8.678 (0.717)	12.495 (1.192)
1978	-0.285 (0.474)	4.462 (0.514)	9.100 (0.722)	10.893 (1.035)
1979	-1.210 (0.479)	3.218 (0.440)	7.252 (0.552)	10.377 (0.819)
1980	-1.262 (0.479)	2.639 (0.452)	7.927 (0.553)	10.048 (0.684)
1981	-1.742 (0.502)	2.920 (0.446)	8.353 (0.606)	11.129 (0.769)
1982	-2.605 (0.511)	2.744 (0.435)	8.872 (0.646)	10.983 (0.798)
1983	-2.142 (0.495)	2.125 (0.452)	6.913 (0.546)	9.599 (0.721)
1984	-2.899 (0.522)	1.822 (0.439)	7.004 (0.548)	10.493 (0.755)
1985	-3.269 (0.506)	2.272 (0.440)	6.088 (0.505)	8.652 (0.588)
1986	-3.539 (0.499)	1.810 (0.456)	5.398 (0.481)	8.029 (0.569)
1987	-3.885 (0.512)	0.984 (0.451)	5.378 (0.493)	8.469 (0.605)
1988	-3.327 (0.530)	1.069 (0.461)	5.507 (0.498)	8.327 (0.600)
1989	-3.927 (0.530)	1.685 (0.460)	5.719 (0.519)	9.007 (0.645)
1990	-3.358 (0.527)	0.973 (0.467)	4.723 (0.488)	8.442 (0.620)
1991	-3.530 (0.514)	0.812 (0.469)	4.536 (0.498)	7.926 (0.587)
1992	-4.189 (0.493)	-0.040 (0.470)	4.217 (0.485)	7.332 (0.553)
1993	-3.783 (0.489)	-0.232 (0.467)	3.441 (0.490)	6.254 (0.521)
1994	-4.898 (0.484)	-0.008 (0.470)	2.884 (0.484)	5.973 (0.538)
1995	-4.702 (0.506)	0.417 (0.482)	3.776 (0.485)	6.266 (0.530)
1996	-4.130 (0.496)	0.891 (0.477)	4.194 (0.493)	7.064 (0.583)
1997	-5.130 (0.501)	-0.438 (0.471)	4.340 (0.492)	6.716 (0.569)
1998	-4.859 (0.485)	-0.794 (0.467)	3.323 (0.478)	6.444 (0.576)
1999	-5.689 (0.501)	-1.544 (0.462)	3.658 (0.494)	7.183 (0.595)
2000	-4.818 (0.503)	-0.569 (0.460)	3.526 (0.514)	7.458 (0.621)
2001	-6.073 (0.513)	-1.050 (0.472)	3.436 (0.493)	7.487 (0.624)
2002	-8.105 (0.544)	-1.978 (0.469)	3.141 (0.507)	8.282 (0.660)
2003	-7.376 (0.511)	-2.158 (0.492)	3.107 (0.519)	9.102 (0.717)
2004	-7.605 (0.539)	-1.901 (0.471)	3.250 (0.532)	7.961 (0.671)
2005	-6.704 (0.517)	-2.564 (0.481)	2.013 (0.514)	8.232 (0.669)
2006	-6.521 (0.517)	-3.158 (0.474)	2.204 (0.539)	8.255 (0.679)
2007	-7.693 (0.554)	-2.787 (0.482)	2.459 (0.526)	7.665 (0.648)
2008	-8.204 (0.575)	-2.437 (0.483)	2.297 (0.534)	7.104 (0.609)
2009	-7.514 (0.555)	-2.573 (0.496)	2.320 (0.538)	7.685 (0.650)
2010	-7.862 (0.574)	-2.403 (0.519)	2.209 (0.548)	8.006 (0.689)

Table 10: Item difficulty posterior estimates and standard deviations for the PTS State item. These parameters are displayed visually in the main document.

t	$Pr(\alpha_{t-1,1} > \alpha_{t,1})$	$Pr(\alpha_{t-1,2} > \alpha_{t,2})$	$Pr(\alpha_{t-1,3} > \alpha_{t,3})$	$Pr(\alpha_{t-1,4} > \alpha_{t,4})$
1976				
1977	0.416	0.702	0.552	0.424
1978	0.570	0.127	0.299	0.780
1979	0.929	0.990	0.766	0.618
1980	0.339	0.798	0.613	0.312
1981	0.578	0.334	0.393	0.477
1982	0.891	0.558	0.250	0.691
1983	0.264	0.793	0.985	0.882
1984	0.828	0.701	0.510	0.139
1985	0.660	0.289	0.870	0.972
1986	0.676	0.700	0.762	0.834
1987	0.721	0.874	0.554	0.228
1988	0.209	0.418	0.429	0.565
1989	0.800	0.213	0.401	0.250
1990	0.182	0.867	0.896	0.721
1991	0.509	0.643	0.582	0.756
1992	0.853	0.908	0.678	0.759
1993	0.345	0.566	0.869	0.908
1994	0.962	0.317	0.818	0.612
1995	0.395	0.272	0.108	0.343
1996	0.224	0.245	0.236	0.208
1997	0.957	0.987	0.393	0.474
1998	0.473	0.587	0.908	0.721
1999	0.772	0.908	0.410	0.236
2000	0.082	0.107	0.483	0.420
2001	0.966	0.788	0.568	0.499
2002	0.996	0.928	0.697	0.226
2003	0.157	0.632	0.528	0.197
2004	0.731	0.265	0.447	0.835
2005	0.215	0.852	0.895	0.371
2006	0.513	0.817	0.305	0.509
2007	0.883	0.347	0.422	0.782
2008	0.712	0.205	0.692	0.798
2009	0.163	0.620	0.541	0.216
2010	0.665	0.407	0.598	0.313

Table 11: The probability that one PTS State cut point is different than the one preceding it in time is displayed here as: $Pr(\alpha_{t-1,k} > \alpha_{t,k})$. These inequality tests are reversed relative to CIRI because the ordering of the original scales are reversed themselves so the PTS cut points should decrease over time whereas the CIRI cut-points should increase. Note that there are many yearly differences, which are highly probably, which corroborates the non-zero linear trend for this as displayed in Table 3 above.

t	$\alpha_{t,1}$	$\alpha_{t,2}$	$\alpha_{t,3}$	$\alpha_{t,4}$
1985	-1.812 (0.298)	-0.110 (0.259)	2.519 (0.284)	4.460 (0.381)
1986	-2.059 (0.316)	0.324 (0.259)	2.792 (0.288)	4.379 (0.391)
1987	-2.049 (0.307)	-0.436 (0.264)	2.505 (0.280)	4.671 (0.451)
1988	-2.496 (0.323)	-0.599 (0.263)	2.187 (0.275)	5.097 (0.487)
1989	-2.594 (0.345)	-0.713 (0.279)	2.176 (0.283)	4.727 (0.440)
1990	-2.447 (0.330)	-0.659 (0.280)	1.745 (0.275)	4.086 (0.385)
1991	-2.405 (0.320)	-0.533 (0.277)	1.659 (0.273)	4.312 (0.411)
1992	-2.650 (0.309)	-0.455 (0.268)	1.902 (0.277)	4.118 (0.392)
1993	-3.015 (0.319)	-0.643 (0.266)	1.816 (0.276)	3.462 (0.338)
1994	-4.354 (0.392)	-0.727 (0.274)	1.967 (0.278)	3.596 (0.351)
1995	-3.874 (0.369)	-0.347 (0.268)	2.073 (0.289)	3.544 (0.362)
1996	-3.876 (0.358)	-0.600 (0.265)	1.642 (0.278)	3.282 (0.347)
1997	-4.738 (0.423)	-1.032 (0.274)	1.652 (0.282)	3.469 (0.356)
1998	-4.805 (0.428)	-1.004 (0.269)	1.220 (0.267)	3.422 (0.356)
1999	-4.664 (0.424)	-1.136 (0.272)	1.516 (0.289)	3.138 (0.369)

Table 12: Item difficulty posterior estimates and standard deviations for the Hathaway torture item. These parameters are displayed visually in the main document.

t	$Pr(\alpha_{t-1,1} > \alpha_{t,1})$	$Pr(\alpha_{t-1,2} > \alpha_{t,2})$	$Pr(\alpha_{t-1,3} > \alpha_{t,3})$	$Pr(\alpha_{t-1,4} > \alpha_{t,4})$
1985				
1986	0.712	0.119	0.249	0.561
1987	0.490	0.980	0.764	0.312
1988	0.844	0.668	0.794	0.261
1989	0.582	0.617	0.510	0.712
1990	0.375	0.444	0.864	0.864
1991	0.463	0.376	0.588	0.344
1992	0.708	0.420	0.267	0.635
1993	0.795	0.693	0.587	0.897
1994	0.996	0.585	0.352	0.391
1995	0.186	0.160	0.396	0.540
1996	0.502	0.753	0.858	0.699
1997	0.940	0.874	0.489	0.356
1998	0.545	0.472	0.866	0.539
1999	0.408	0.636	0.225	0.711

Table 13: The probability that one Hathaway torture cut point is different than the one preceding it in time is displayed here as: $Pr(\alpha_{t-1,k} > \alpha_{t,k})$. These inequality tests are reversed relative to CIRI because the ordering of the original scales are reversed themselves so the Hathaway torture cut points should decrease over time whereas the CIRI cut-points should increase. Note that there are many yearly differences, which are highly probably, which corroborates the non-zero linear trend for this as displayed in Table 3 above.

t	$\alpha_{t,1}$	$\alpha_{t,2}$	$\alpha_{t,3}$	$\alpha_{t,4}$	$\alpha_{t,5}$
1995	-0.860 (0.226)	-0.771 (0.222)	-0.101 (0.222)	0.663 (0.223)	1.162 (0.237)
1996	-1.170 (0.210)	-1.079 (0.207)	-0.392 (0.200)	0.484 (0.200)	1.437 (0.227)
1997	-1.105 (0.215)	-0.972 (0.211)	-0.483 (0.204)	0.465 (0.209)	1.177 (0.224)
1998	-1.081 (0.222)	-0.987 (0.219)	-0.349 (0.212)	0.334 (0.214)	0.942 (0.227)
1999	-1.243 (0.236)	-1.198 (0.234)	-0.276 (0.217)	0.561 (0.223)	1.236 (0.235)
2000	-1.899 (0.250)	-1.736 (0.243)	-1.049 (0.218)	-0.063 (0.214)	1.193 (0.237)
2001	-1.761 (0.242)	-1.664 (0.235)	-0.687 (0.211)	-0.007 (0.206)	0.879 (0.225)
2002	-1.527 (0.222)	-1.351 (0.216)	-0.630 (0.208)	0.191 (0.207)	1.167 (0.223)
2003	-1.229 (0.207)	-1.068 (0.203)	-0.362 (0.194)	0.173 (0.203)	1.228 (0.227)
2004	-0.696 (0.206)	-0.592 (0.205)	0.007 (0.201)	0.417 (0.203)	1.290 (0.230)
2005	-0.536 (0.204)	-0.322 (0.208)	0.245 (0.211)	0.528 (0.214)	1.420 (0.243)

Table 14: Item difficulty posterior estimates and standard deviations for the ITT torture item. These parameters are displayed visually in the main document.

t	$Pr(\alpha_{t-1,1} > \alpha_{t,1})$	$Pr(\alpha_{t-1,2} > \alpha_{t,2})$	$Pr(\alpha_{t-1,3} > \alpha_{t,3})$	$Pr(\alpha_{t-1,4} > \alpha_{t,4})$	$Pr(\alpha_{t-1,5} > \alpha_{t,5})$
1995					
1996	0.845	0.847	0.836	0.726	0.200
1997	0.415	0.357	0.625	0.528	0.791
1998	0.471	0.518	0.323	0.673	0.769
1999	0.694	0.746	0.405	0.231	0.184
2000	0.972	0.945	0.994	0.978	0.552
2001	0.345	0.420	0.116	0.424	0.831
2002	0.238	0.162	0.421	0.252	0.185
2003	0.161	0.171	0.173	0.523	0.426
2004	0.035	0.050	0.094	0.197	0.425
2005	0.292	0.179	0.207	0.354	0.351

Table 15: The probability that one ITT torture cut point is different than the one preceding it in time is displayed here as: $Pr(\alpha_{t-1,k} > \alpha_{t,k})$. These inequality tests are reversed relative to CIRI because the ordering of the original scales are reversed themselves so the ITT cut points should decrease over time whereas the CIRI cut-points should increase. Note that there are some yearly differences, which are highly probably even though the linear trend for this item is flat and not probabilistically different than 0 as displayed in Table 3 above.

H Item Discrimination Paramters

	Dynamic Standard	Constant Standard	$Pr(\beta_j^{dynamic} > \beta_j^{constant})$
CIRI Physical Integrity Data			
political imprisonment	1.655 (0.091)	1.584 (0.102)	0.698
torture	2.020 (0.110)	1.870 (0.119)	0.823
extrajudicial killing	2.431 (0.134)	2.401 (0.155)	0.561
disappearance	2.287 (0.132)	2.237 (0.151)	0.596
Hathaway Torture Data			
torture	1.843 (0.110)	1.742 (0.116)	0.737
Ill-Treatment and Torture			
torture	0.976 (0.071)	0.978 (0.079)	0.495
Political Terror Scale			
State	4.494 (0.269)	4.265 (0.270)	0.576
Amnesty	3.220 (0.177)	3.167 (0.200)	0.727
Harff and Gurr			
massive repression	6.379 (0.587)	5.482 (0.531)	0.871
PITF			
genocide and politicide	4.313 (0.343)	3.247 (0.273)	0.992
Rummel			
genocide and democide	4.452 (0.318)	4.147 (0.317)	0.752
UCDP			
killing	2.284 (0.180)	2.347 (0.202)	0.407
WHPSI			
executions	0.965 (0.078)	0.895 (0.076)	0.739

Table 16: Posterior estimates and 95% credible intervals for item discrimination parameters β from the dynamic standard model and constant standard model. In alternative parameterizations using a normal prior, none of these parameters overlap with 0. The item discrimination parameters for each model allow assessment of the information value of each indicator. All of the items discriminated well with respect to the latent variable in both models though more so in the dynamic standard model relative to the constant standard model for many of the items. This systemic difference is evidence of the improved fit of the dynamic standard model relative to the constant standard model. The wide range of item discrimination parameters for both models provides substantial evidence that these models both improve on the additive scales used throughout most of the human rights literature as first reported by Schnakenberg and Fariss (2013).

I Deviance Information Criterion (DIC)

The Deviance Information Criterion (DIC) is a method useful for comparing the relative fit of item response theory models because the model with the smallest DIC is expected to have the greatest out of sample predictive power (Spiegelhalter et al., 2002). The DIC is also useful for comparing the models in this article because it penalizes more complex models so that the more parsimonious model is favored, all else equal (Gelman et al., 2003). For a given factor of parameters Ψ, the deviance is given by $D(y,\Psi) = -2\log(\mathscr{L}(y|\Psi))$ where $\mathscr{L}(y|\Psi)$ is the likelihood function of the model. Other commonly used information criteria use the number of parameters as an argument, but in a hierarchical context the number of parameters can be difficult to quantify. The DIC uses the *effective number of parameters* which is $pD = \overline{D}(y) - \hat{D}(y,\hat{\Psi})$ where $\overline{D}(y)$ is the posterior mean of the deviance and $\hat{D}(y,\hat{\Psi})$ is the deviance estimates using the posterior mean of the parameters, $\hat{\Psi}$. The DIC is $DIC = 2\overline{D}(y) - \hat{D}(y,\hat{\Psi})$. The differences obtained in comparing the two models is several thousand in favor of the dynamic standard model as displayed in Table 17.

DIC	**Constant**	**Dynamic**
Mean deviance	52492	50587
penalty	2535	3119
Penalized deviance	**55027**	**53706**

Table 17: Deviance Information Criterion statistics for two models. The model with the dynamic standard of accountability (time varying difficulty cut-points per) performs better (smaller deviance) than the model with the constant standard of accountability (constant difficulty cut-points).

J Posterior Predictive Checks: Additional Table

Repression Variable	Proportion
CIRI Physical Integrity Data	
political imprisonment	0.477
torture	0.632
extrajudicial killing	0.525
disappearance	0.477
Hathaway Torture Data	
torture	0.601
Ill-Treatment and Torture	
torture	0.544
Political Terror Scale	
State	0.493
Amnesty	0.622
Harff and Gurr	
massive repression	0.835
PITF	
genocide and politicide	0.832
Rummel	
genocide and democide	0.618
UCDP	
killing	0.542
WHPSI	
executions	0.632
Average	0.602

Table 18: The column values measure the proportion of country-year observations that have a smaller sum of squared deviation generated by comparing the observed item and predicted item for the dynamic standard model compared to the constant standard model. The dynamic standard model does a better job at predicting the repression variable compared to predictions generated from the constant standard model. This information is displayed in a figure in the main document. 2000 posterior draws were used to generate these statistics, they are therefore highly accurate estimates.

K Analyzing the Standards-Based Repression Variables

The procedure I describe in this section is useful for analyzing the original standards-based repression variables included in this article. I estimate an ordered logistic regression model for each of the original standards-based repression variables. I also run a model with the combined CIRI indicators, which is simply the original CIRI additive scale.

For these models, I regress the ordered repression variable on (1) the lagged value of the original ordered variable itself, (2) the lagged value of the latent repression variables from the constant standard model, and (3) the lagged value of the latent repression variables from the dynamic standard model (the numbers denote columns in Table 19). To compare these alternative models I generate a statistic known as model deviance. Just like a sum of squared deviance statistic, smaller values of model deviance indicate a better fitting model. For the ordered logistic regression models using the latent repression estimates, I estimate 1000 regressions by taking a draw from the posterior mean and standard deviation for each country-year of the latent variable before I estimate the regressions. This procedure allows me to incorporate uncertainty into the model deviance statistics.

Temporal bias exists in both the ordered data taken from the standards-based sources and consequently, the latent variable estimates from the constant standard model, which does not account for the changing standard of accountability. It is therefore unsurprising that the lagged latent variable from the constant standard model generates lower model deviance statistics when paired with many of the original ordered repression variables (see Section 10.10 of the appendix).

There is a straight forward correction to account for this temporal bias: simply include (1) the lagged latent variable from the dynamic standard model, (2) an indicator for time t, where $t = 1, \ldots, T$, and (3) the interaction of the index of time and the lagged repression variable. Simply estimate the following model specification in R: `polr(as.factor(y) ~ theta_t-1 + t + t*theta_t-1 + ...)`, where `y` is any of the original ordered standards-based variables, `t` is the time index and `theta_t-1` is the lagged latent variable from the dynamic standard model. The same model can be estimated in Stata using the **ologit** command with the same variables. This specification accounts for the bias in the original ordered variables but only for those models using the lagged variable from the dynamic standard model.

This specification is effective because the temporal bias no longer exists in the latent variable estimated from the dynamic standard model, which is interacted with the time index. The interaction is necessary though, because temporal bias does exist in the original standards-based dependent variables. The specification must include the interaction of the lagged latent repression variable with the index of time when analyzing the original ordered data. Intuitively, the value of the standards-based repression variable is conditional on the value of the lagged variable but, as I demonstrated above, this conditional relationship changes over time. The interaction specification captures this idea.

The models with the lagged latent variable from the dynamic standard model interacted with the index of time are the best fitting models for nearly all the tests presented in this section. The models using the lagged latent variable form the dynamic standard model are always better at predicting the original repression variables relative to the lagged latent variable from the constant standard model. This specification represents a method to continue to model ordered repression variables, which is especially useful for analyzing one of the disaggregated CIRI variables.

Dependent Variable	**Lagged Repression Variables**		
	t*Y_{t-1}	t*Constant Standard$_{t-1}$	t*Dynamic Standard$_{t-1}$
CIRI Physical Integrity			
Additive Scale	13088	12944 [12865, 13032]	**12222 [12142, 12315]**
political imprisonment	**5846**	7067 [7030, 7104]	6895 [6851, 6935]
torture	6234	6095 [6049, 6143]	**5792 [5744, 5846]**
extrajudicial killing	6069	5867 [5820, 5913]	**5570 [5515, 5620]**
disappearance	4213	4187 [4151, 4227]	**3995 [3953 4033]**
Hathaway Torture			
torture	**4241**	4668 [4634, 4700]	4490 [4453, 4527]
Ill-Treatment and Torture			
torture	**3116**	3495 [3479, 3511]	3467 [3450, 3483]
Political Terror Scale			
State	8758	8296 [8211, 8384]	**7428 [7321, 7530]**
Amnesty	8102	8143 [8070, 8219]	**7502 [7423, 7576]**

Table 19: Model deviance statistics from ordered logistic regression models. These models differ from those estimated using the event-based binary data because of the inclusion of t and the interaction of t with the lagged repression variable. This interaction specification accounts for the bias in the original ordered variables but only for those models using the lagged variable from the dynamic standard model. Each row represents the model deviance statistics from three ordered logistic regression models estimated for comparison. Smaller values across rows indicate a better fitting model. The best fitting model is in bold. These statistics are not standardized and should only be compared across rows.

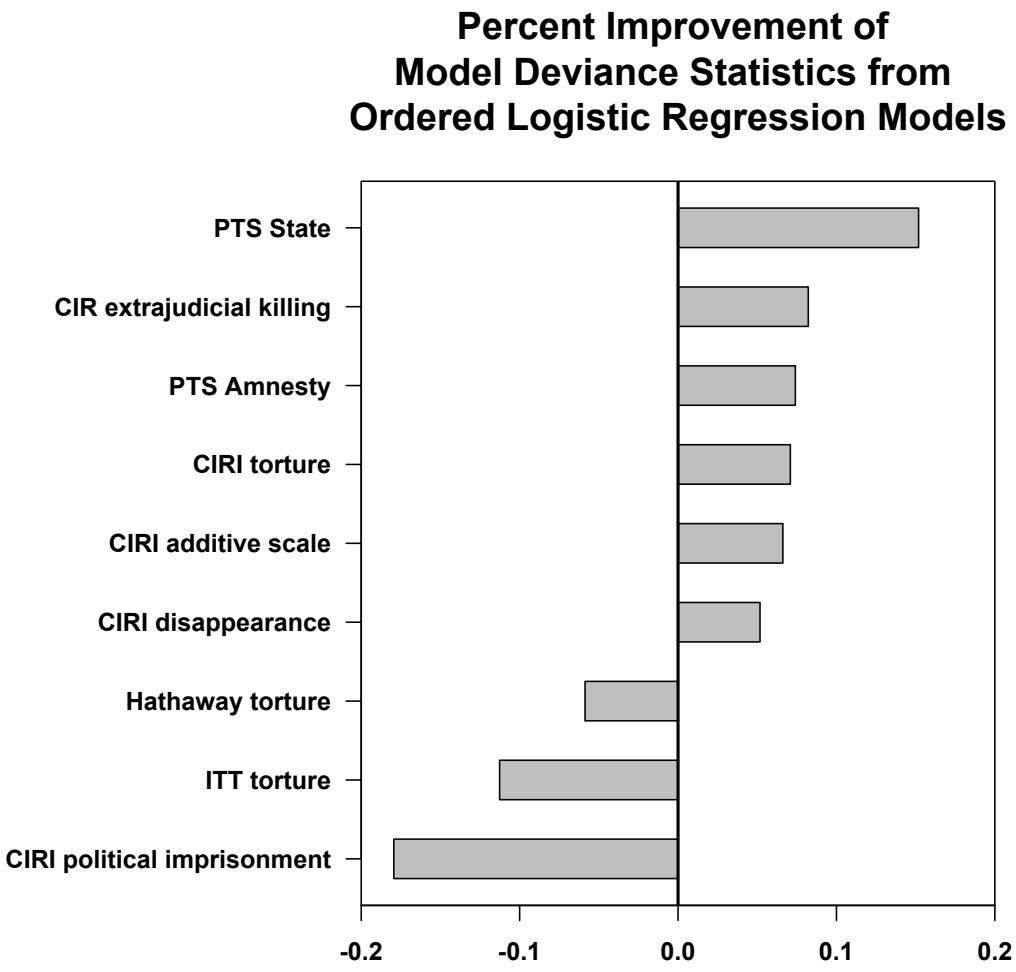

Figure 14: Visual representation of the percentage improvement for the model deviance statistics from the based line models (original standards based data) in Table 19 and the models using the latent variable estimates generated from the dynamic standard model.

Dependent Variable	Lagged Repression Variables		
	Y_{t-1}	*Constant Standard*$_{t-1}$	*Dynamic Standard*$_{t-1}$
CIRI Physical Integrity			
Additive Scale	13093	**12950 [12871, 13026]**	12958 [12883, 13035]
political imprisonment	**5862**	7221 [7182, 7256]	6916 [6874, 6958]
torture	6278	**6212 [6167, 6258]**	6521 [6478, 6564]
extrajudicial killing	6077	**5882 [5834, 5927]**	6007 [5959, 6055]
disappearance	4243	4246 [4206, 4285]	**4031 [3990, 4070]**
Hathaway Torture			
torture	**4241**	4677 [4643, 4709]	4575 [4539, 4611]
Ill-Treatment and Torture			
torture	**3123**	3506 [3492, 3522]	3470 [3453, 3486]
Political Terror Scale			
State	8760	**8379 [8292, 8467]**	9343 [9258, 9426]
Amnesty	8119	8266 [8196, 8331]	**7849 [7770, 7922]**

Table 20: Model deviance statistics from binary ordered logistic regression models. Each row represents the model deviance statistics from three logistic regression models estimated for comparison. Smaller values across rows indicate a better fitting model. The best fitting model is in bold. These statistics are not standardized and should only be compared across rows and also between rows in this table and the one presented in the main document. Keep in mind that these statistics are different from those presented in the main document because they are not derived from models that include an interaction with the lagged repression variable and the index for time. The interaction term is necessary to account for the changing standard of accountability that affects the reports from which the standards-based variables are derived. When the interaction term is included in the model, as presented in the main document, the estimates in the third column are always smaller than the estimates in the second column.

L Analyzing the Event-Based Repression Variables

The procedure I describe in this section is useful for analyzing the original event-based repression variables included in this article. I specify binary logistic regression models since the event-based data are all binary. I regress each binary variable on (1) yearly dummy variables, (2) cubic polynomials of time since the last event (3) the lagged value of the original variable itself, (4) the lagged value of the latent repression variable from the constant standard model, and (5) the lagged value of the latent repression variable from the dynamic standard model (the numbers denote columns in Table 21).

To compare these models I generate a statistic known as the area under the receiver operator curve or AUC for short. A value of 1 for this statistic indicates that the model perfectly predicts the outcome; a value of 0.5 indicates the model predicts the data no better than chance. For the binary logistic regression models that use the latent repression estimates, I estimate 1000 regressions by taking a draw from the posterior mean and standard deviation for each country-year to incorporate uncertainty into the resulting AUC statistics. I report only a single AUC statistic for the other models.

The estimates generated from the models that include the lagged dynamic standard variables outperform the alternatives for all of the event-based variables (see Table 21). This is an important result because it demonstrates that bias in the constant standard model reduces the ability of repression estimates from this model to predict future event-based outcomes.

Analysts wishing to model the binary event-based repression variables can now use the lagged version of the repression estimates generated from the dynamic standard model. This model specification represents an alternative to the current practice of specifying a duration dependent binary variable with temporal dummy variables, natural cubic splines, or a temporal polynomial (e.g., Beck, Katz and Tucker, 1998; Carter and Signorino, 2010). The model specification that includes the lagged version of the repression estimates generated from the dynamic standard model out perform these alternative specifications. To estimate this model, simply use the following specification in R: `glm(y ~ theta_t-1 + ..., binomial(link = "logit"))`, where `y` is any of the original binary event-based variables and `theta_t-1` is the lagged latent variable from the dynamic standard model. The same model can be estimated in Stata using the **logit** command with the same lagged variable.

Dependent Variable	Temporal Controls		Lagged Repression Variables		
	t dummies	t^1, t^2, t^3	Y_{t-1}	*Constant Standard*$_{t-1}$	*Dynamic Standard*$_{t-1}$
Harff and Gurr massive repression	0.621	0.928	0.941	0.961 [0.957, 0.964]	**0.981** [0.978, 0.984]
PITF genocide and politicide	0.707	0.937	0.933	0.947 [0.943, 0.950]	**0.975** [0.973, 0.978]
Rummel genocide and democide	0.544	0.951	0.967	0.956 [0.954 0.958]	**0.974** [0.972, 0.976]
UCDP killing	0.598	0.843	0.786	0.890 [0.886, 0.895]	**0.918** [0.913, 0.922]
WHPSI executions	0.586	0.752	0.661	0.761 [0.751, 0.770]	**0.779** [0.769, 0.788]

Table 21: AUC statistics from binary logistic regression models. Each row represents the AUC statistics from five logistic regression models estimated for comparison. A value of 1 for this statistic indicates that the model perfectly predicts the outcome; a value of 0.5 indicates the model predicts the data no better than chance. AUC statistics in bold font represent the best fitting model in each row. All of the lagged repression variables generated from the dynamic standard model out perform the other lagged variables they were compared against. This includes the additional models that use standard temporal controls. The first column presents AUC statistics generated from a model that regresses the binary dependent variable on yearly dummy variables, which are coded 1 for year t and 0 otherwise. The second column presents AUC statistics generated from a model that regresses the binary dependent variable on the cubic polynomial of the duration between spells (i.e., the period of time between instances when the dependent variable is coded 1). See Beck, Katz and Tucker (1998) and Carter and Signorino (2010) for a discussion of these techniques. The lagged repression variables generated from the dynamic standard model out perform every alternative.

M Method for Incorporating Uncertainty into a Model

Schnakenberg and Fariss (2013) describe a technique, which is designed to incorporate measurement uncertainty into any model that includes a latent variable on the right hand side of a regression equation. The procedure is to create m datasets, which can be as low as 5 or 10 (Mislevy, 1991). The datasets are constructed using different draws from the posterior distribution of the latent variable and then combined using the Rubin (1987) formulas, where the point estimate for each parameter is the mean from the m estimates, and the standard error is $\sqrt{\frac{1}{m}\sum_k^m s_k^2 + \left(1+\frac{1}{m}\right)\sigma_\beta^2}$ where s_k^2 is the standard error from dataset k, and σ_β^2 is the variance in the regression coefficients between datasets. In words, the standard error is the average standard error from each model, plus the variance in the regression coefficients times a correction factor for $m < \infty$. This is the same procedure used for multiple imputation in the political science community (King et al., 2001).

N Information About The Convention Against Torture Model Specifications

I use the method described in above in Section M to incorporate uncertainty in the following regression models. These two models are used to generate the coefficients that measure the association between one the competing latent variables respectively and ratification of the UN Convention Against Torture. Both models also include additional control variables. Note that the models only include data from 1976 until 2005 because of data availability. This exclusion of the observations from 2006-2010 reduces the chance that coefficients in the competing models will be different since the temporal bias increases with respect to time.

Linear Regression Coefficients from the Main Text

Table 22: Linear Model with Dependent Variable from the Dynamic Standard Model

Variables	Coefficients	Std. Errors
Intercept	0.186	0.085
DV_{t-1}	0.902	0.007
Convention Against Torture	0.022	0.014
$Polity_{t-1}$	0.007	0.001
$ln(Population_{t-1})$	-0.034	0.005
$ln(GDP per\ capita_{t-1})$	0.045	0.007

Table 23: Linear Model with Dependent Variable from the Constant Standard Model

Variables	Coefficients	Std. Errors
Intercept	0.123	0.089
DV_{t-1}	0.893	0.007
Convention Against Torture	-0.036	0.015
$Polity_{t-1}$	0.007	0.001
$ln(Population_{t-1})$	-0.035	0.005
$ln(GDP per\ capita_{t-1})$	0.051	0.007

In addition to the sign flip, the difference between the coefficients for the Convention Against Torture binary variable generated in the two competing modes is 0.0593 ($p < 0.004$). The p-value for this

difference is simply based on the following $Z-score$: $\frac{\beta_{dynamic}-\beta_{constant}}{\sqrt{SE(\beta_{dynamic})^2+SE(\beta_{constant})^2}}$. Note further that, none of the other coefficients are different form one another.

Additional Linear Regression Coefficients not from the Main Text

Table 24: Linear Model with Dependent Variable from the Dynamic Standard Model

Variables	Coefficients	Std. Errors
Intercept	0.267	0.084
DV_{t-1}	0.871	0.009
Convention Against Torture	0.015	0.015
$ln(Population_{t-1})$	-0.033	0.005
$Population\ growth_{t-1}$	-0.010	0.005
$ln(GDP per\ capita_{t-1})$	0.044	0.005
$GDP\ per\ capita\ growth_{t-1}$	0.000	0.001
$Polity_{t-1}$	0.005	0.001
$International\ War_{t-1}$	-0.009	0.027
$Civil\ War_{t-1}$	-0.155	0.022
$Military\ Regime_{t-1}$	-0.018	0.017
$British\ Colonial\ Legacy$	0.001	0.015

Table 25: Linear Model with Dependent Variable from the Constant Standard Model

Variables	Coefficients	Std. Errors
Intercept	0.176	0.086
DV_{t-1}	0.860	0.008
Convention Against Torture	-0.048	0.016
$ln(Population_{t-1})$	-0.030	0.005
$Population\ growth_{t-1}$	-0.008	0.005
$ln(GDP per\ capita_{t-1})$	0.046	0.006
$GDP\ per\ capita\ growth_{t-1}$	0.002	0.001
$Polity_{t-1}$	0.006	0.001
$International\ War_{t-1}$	-0.023	0.028
$Civil\ War_{t-1}$	-0.205	0.022
$Military\ Regime_{t-1}$	-0.026	0.017
$British\ Colonial\ Legacy$	0.001	0.015

Note that in this specification, the sign flip still occurs and the difference between the coefficients for the Convention Against Torture binary variable generated in the two competing modes is 0.0633 $(p < 0.004)$. And, though the standard error on the CAT coefficient in the model using the DV from the

dynamic standard model is the same, the size of the effect has reduced, which decreases the statistical significance between this coefficient and 0. Nonetheless, the size of the difference between the two coefficient estimates is the same magnitude and level of significance. These differences are seen in the side-by-see plot of the difference in Figure fig:CompareDIFF. Note the variable for this specification are taken from Poe, Rost and Carey (2006), which is one of the same specifications as found in Poe and Tate (1994) and Poe, Tate and Keith (1999).

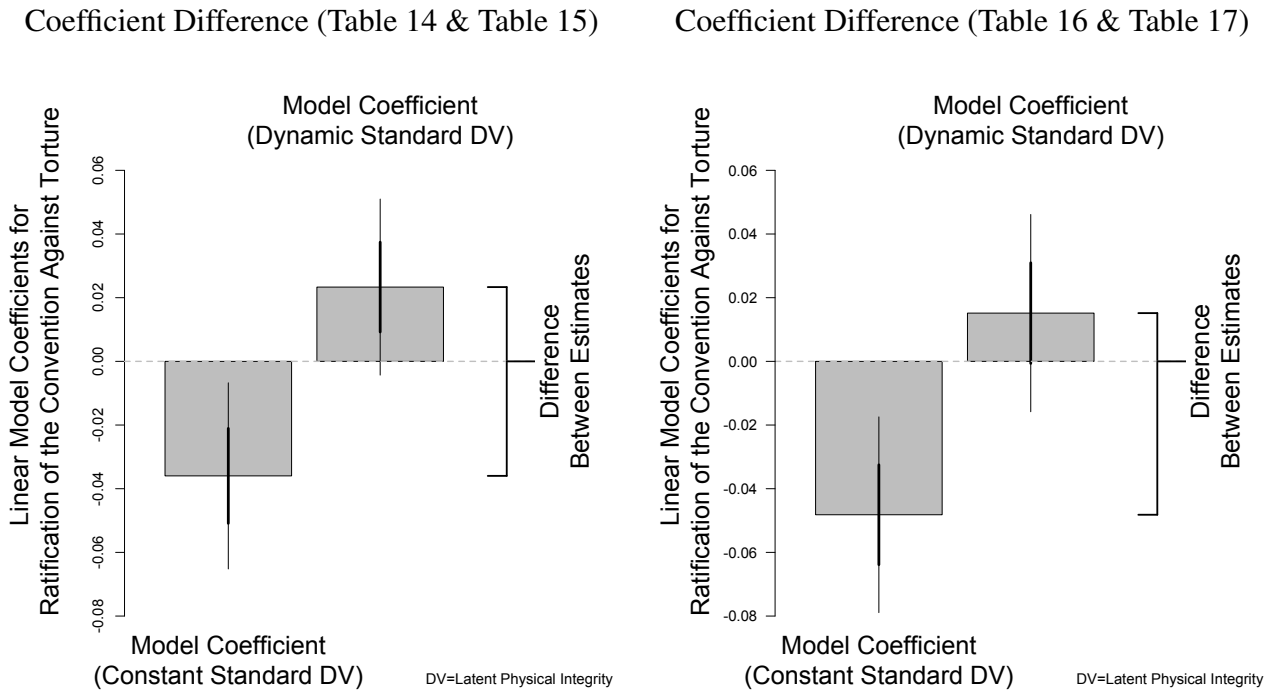

Figure 15: Estimated coefficient for CAT (the UN Convention Against Torture) ratification from the linear model using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error of the coefficient. The thin lines represent 2 $\pm$ the standard error of the coefficient. The difference between the coefficients is statistically significant ($p < 0.004$) in both model specifications.

O Trace Plots

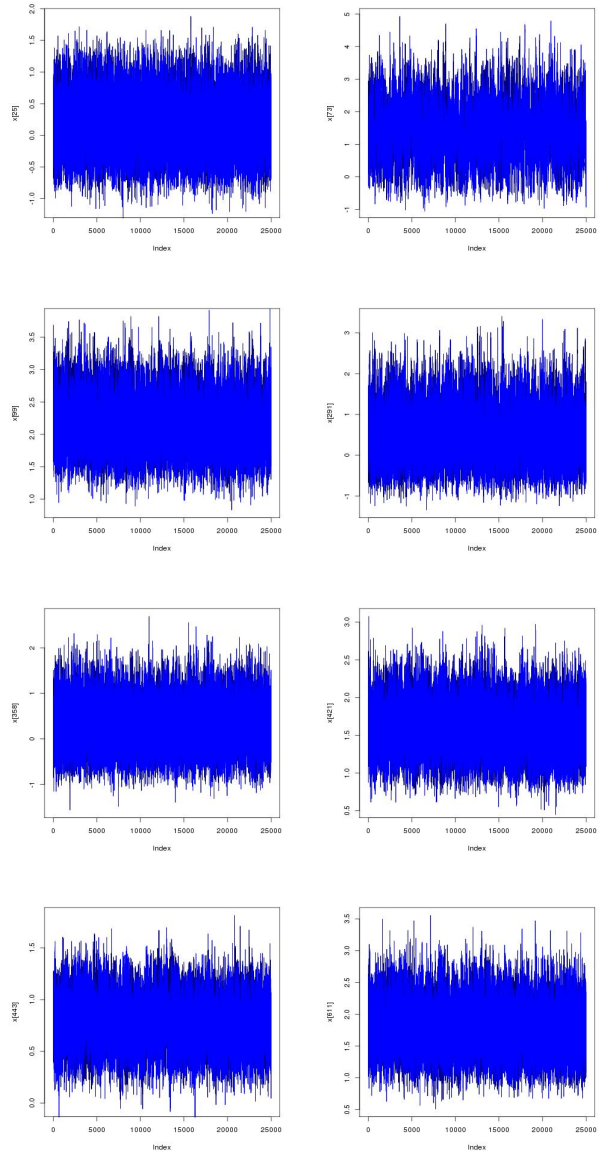

Figure 16: Trace plots for eight country-year latent variable estimates. These example plots visualize the converged MCMC chains. Plots of the other parameters follow a similar pattern and are available form the author.

P Latent Variable Model Extensions

I have estimated versions of the model that use a multivariate normal prior and multivariate Wishart prior for the item parameters (Gelman and Hill, 2007). This allows some correlation among the items that is not due to the latent variable itself, which relaxes the local independence assumption, which I describe in detail in the article. Note that the main inference presented in the article is robust to these extensions. Specifically, the latent variables estimated from these model are essentially the same. I leave further exploration of the relationship between the items to future research. I believe that this is a promising avenue for future research.

I have also attempted to estimate a 2-dimensional latent variable model instead of the unidimensional version presented throughout the article and appendix. This version of the model does not converge however. The first dimension of the latent variable continues to correlate highly with its single-dimensional counterpart even though most of the other model parameters do not appear to be anywhere close to convergence based on visual inspection of trace plots. Graded response models generally are difficult to estimate in the first place and even more so with multidimensional latent traits as described by Albert and Johnson (1999). However, I believe that this is a promising avenue for future research if additional data directly attributable to another dimension are obtained. For example it should be possible to estimate a multidimensional human rights model in which the first dimension is based on physical integrity items and the second dimension is based on some other set of human rights indicators such as the empowerment variables from the CIRI human rights project. Allowing these two dimensions to inform each other in such an extended model would likely lead to even more informative estimates of both latent variables. A democracy dimension could also be estimated jointly using existing indicators of that construct. I leave these suggestions to future work and make a note about the need to reconsider existing democracy scales with respect to changing standards and judgment coding in the conclusion of the article.

Q Bugs Model Files

Note that these .bug files are not as condensed as the could be but they are probably more informative this expanded form. Neither the results nor the computation speed are contingent on the compactness of these model files. Recall that each of these models are estimated with two MCMC chains, which are run 100,000 iterations using JAGS (Plummer, 2010) on the Gordon Supercomputer (Sinkovits et al., 2011). The first 50,000 iterations were thrown away as burn-in and the rest were used for inference. Diagnostics all suggest convergence (Geweke, 1992; Heidelberger and Welch, 1981, 1983; Gelman and Rubin, 1992). Finally, note that I've changed the latent variable θ to x in the model file for ease of viewing.

Bugs Model Code for Dynamic Standard Model

```
model{
 for(i in 1:n){# n is the number of obs

# CIRI items, DISAP, KILL, POLPRIS, TORT  standards based data
   for(item in 1:4){
    logit(Z[i, item, 1])  <- alpha3[item , 1, time[i]] - beta3[item] * x[i]
    logit(Z[i, item, 2])  <- alpha3[item , 2, time[i]] - beta3[item] * x[i]
    Pi[i, item, 1] <- Z[i, item, 1]
    Pi[i, item, 2] <- Z[i, item, 2] - Z[i, item, 1]
    Pi[i, item, 3] <- 1 - Z[i, item, 2]
    y[i, item] ~ dcat(Pi[i, item, 1:3])
   }

# PTS Amnesty standards based data
    logit(Z[i, 5, 1])  <- alpha5[1, 1, time[i]] + beta5[1]*x[i]
    logit(Z[i, 5, 2])  <- alpha5[1, 2, time[i]] + beta5[1]*x[i]
    logit(Z[i, 5, 3])  <- alpha5[1, 3, time[i]] + beta5[1]*x[i]
    logit(Z[i, 5, 4])  <- alpha5[1, 4, time[i]] + beta5[1]*x[i]
    Pi[i, 5, 1] <- Z[i, 5, 1]
    Pi[i, 5, 2] <- Z[i, 5, 2] - Z[i, 5, 1]
    Pi[i, 5, 3] <- Z[i, 5, 3] - Z[i, 5, 2]
    Pi[i, 5, 4] <- Z[i, 5, 4] - Z[i, 5, 3]
    Pi[i, 5, 5] <- 1 - Z[i, 5, 4]
    y[i, 5] ~ dcat(Pi[i, 5, 1:5])

# PTS State standards based data
    logit(Z[i, 6, 1])  <- alpha5[2, 1, time[i]] + beta5[2]*x[i]
    logit(Z[i, 6, 2])  <- alpha5[2, 2, time[i]] + beta5[2]*x[i]
    logit(Z[i, 6, 3])  <- alpha5[2, 3, time[i]] + beta5[2]*x[i]
    logit(Z[i, 6, 4])  <- alpha5[2, 4, time[i]] + beta5[2]*x[i]
    Pi[i, 6, 1] <- Z[i, 6, 1]
    Pi[i, 6, 2] <- Z[i, 6, 2] - Z[i, 6, 1]
    Pi[i, 6, 3] <- Z[i, 6, 3] - Z[i, 6, 2]
    Pi[i, 6, 4] <- Z[i, 6, 4] - Z[i, 6, 3]
    Pi[i, 6, 5] <- 1 - Z[i, 6, 4]
    y[i, 6] ~ dcat(Pi[i, 6, 1:5])

# Hathaway standards based data
    logit(Z[i, 7, 1])  <- alpha5[3, 1, time[i]] + beta5[3]*x[i]
    logit(Z[i, 7, 2])  <- alpha5[3, 2, time[i]] + beta5[3]*x[i]
    logit(Z[i, 7, 3])  <- alpha5[3, 3, time[i]] + beta5[3]*x[i]
    logit(Z[i, 7, 4])  <- alpha5[3, 4, time[i]] + beta5[3]*x[i]
    Pi[i, 7, 1] <- Z[i, 7, 1]
    Pi[i, 7, 2] <- Z[i, 7, 2] - Z[i, 7, 1]
    Pi[i, 7, 3] <- Z[i, 7, 3] - Z[i, 7, 2]
    Pi[i, 7, 4] <- Z[i, 7, 4] - Z[i, 7, 3]
    Pi[i, 7, 5] <- 1 - Z[i, 7, 4]
    y[i, 7] ~ dcat(Pi[i, 7, 1:5])

# ITT
    logit(Z[i, 8, 1])  <- alpha6[1, 1, time[i]] + beta6[1]*x[i]
    logit(Z[i, 8, 2])  <- alpha6[1, 2, time[i]] + beta6[1]*x[i]
    logit(Z[i, 8, 3])  <- alpha6[1, 3, time[i]] + beta6[1]*x[i]
    logit(Z[i, 8, 4])  <- alpha6[1, 4, time[i]] + beta6[1]*x[i]
```

```
    logit(Z[i, 8, 5])  <- alpha6[1, 5, time[i]] + beta6[1]*x[i]
    Pi[i, 8, 1] <- Z[i, 8, 1]
    Pi[i, 8, 2] <- Z[i, 8, 2] - Z[i, 8, 1]
    Pi[i, 8, 3] <- Z[i, 8, 3] - Z[i, 8, 2]
    Pi[i, 8, 4] <- Z[i, 8, 4] - Z[i, 8, 3]
    Pi[i, 8, 5] <- Z[i, 8, 5] - Z[i, 8, 4]
    Pi[i, 8, 6] <- 1 - Z[i, 8, 5]
    y[i, 8] ~ dcat(Pi[i, 8, 1:6])

# Genocide event data
   logit(p[i,1]) <- alpha1[1] - beta1[1]*x[i]
   y[i, 9] ~ dbern(p[i,1])

# Rummel event data
   logit(p[i,2]) <- alpha1[2] - beta1[2]*x[i]
   y[i, 10] ~ dbern(p[i,2])

# Massive Repression data
   logit(p[i,3]) <- alpha1[3] - beta1[3]*x[i]
   y[i, 11] ~ dbern(p[i,3])

# WHPSI killing  event data
   logit(p[i,4]) <- alpha1[4] - beta1[4]*x[i]
   y[i, 12] ~ dbern(p[i,4])

# UPSALA killing  event data
   logit(p[i,5]) <- alpha1[5] - beta1[5]*x[i]
   y[i, 13] ~ dbern(p[i,5])

# redraw latent variable parameter from mu matrix because of unbalanced panels
   x[i] <- mu[country[i], year[i]]
 }

# draw percision for latent variable parameter estimate
 sigma ~ dunif(0,1)
 kappa <- pow(sigma, -1)

# draw dynamic latent variable parameter
 for(c in 1:n.country){
  mu[c, 1] ~ dnorm(0, 1)
  for(t in 2:n.year){ #n.year is number of years
   mu[c, t] ~ dnorm(mu[c, t-1], kappa)
  }
 }

# CIRI model parameters
 for(item3 in 1:4){
  beta3[item3] ~ dgamma(4, 3)
  alpha03[item3, 1, 1] ~ dnorm(0, .25)
  alpha03[item3, 2, 1] ~ dnorm(0, .25)
  alpha3[item3, 1:2, 1] <- sort(alpha03[item3, 1, 1:2])
   for(t in 2:n.year){ #n.year is number of years
    alpha03[item3, 1, t] ~ dnorm(alpha03[item3, 1, t-1], .25)
    alpha03[item3, 2, t] ~ dnorm(alpha03[item3, 2, t-1], .25)
    alpha3[item3, 1:2, t] <- sort(alpha03[item3, 1:2, t])
 }
```

```
}
# PTS and Hathaway model parameters
 for(item5 in 1:3){
  beta5[item5] ~ dgamma(4, 3)
  alpha05[item5, 1, 1] ~ dnorm(0, .25)
  alpha05[item5, 2, 1] ~ dnorm(0, .25)
  alpha05[item5, 3, 1] ~ dnorm(0, .25)
  alpha05[item5, 4, 1] ~ dnorm(0, .25)
  alpha5[item5, 1:4, 1] <- sort(alpha05[item5, 1:4, 1])
   for(t in 2:n.year){ #n.year is number of years
    alpha05[item5, 1, t] ~ dnorm(alpha05[item5, 1, t-1], .25)
    alpha05[item5, 2, t] ~ dnorm(alpha05[item5, 2, t-1], .25)
    alpha05[item5, 3, t] ~ dnorm(alpha05[item5, 3, t-1], .25)
    alpha05[item5, 4, t] ~ dnorm(alpha05[item5, 4, t-1], .25)
    alpha5[item5, 1:4, t] <- sort(alpha05[item5, 1:4, t])
 }
}
 # ITT model parameters
 for(item6 in 1:1){
  beta6[item6] ~ dgamma(4, 3)
  alpha06[item6, 1, 1] ~ dnorm(0, .25)
  alpha06[item6, 2, 1] ~ dnorm(0, .25)
  alpha06[item6, 3, 1] ~ dnorm(0, .25)
  alpha06[item6, 4, 1] ~ dnorm(0, .25)
  alpha06[item6, 5, 1] ~ dnorm(0, .25)
  alpha6[item6, 1:5, 1] <- sort(alpha06[item6, 1:5, 1])
   for(t in 2:n.year){ #n.year is number of years
     alpha06[item6, 1, t] ~ dnorm(alpha06[item6, 1, t-1], .25)
     alpha06[item6, 2, t] ~ dnorm(alpha06[item6, 2, t-1], .25)
     alpha06[item6, 3, t] ~ dnorm(alpha06[item6, 3, t-1], .25)
     alpha06[item6, 4, t] ~ dnorm(alpha06[item6, 4, t-1], .25)
     alpha06[item6, 5, t] ~ dnorm(alpha06[item6, 5, t-1], .25)
     alpha6[item6, 1:5, t] <- sort(alpha06[item6, 1:5, t])
 }
}

# Genocide, Rummel, Massive Repression, UCDP killing and WHPSI execution parameters
 for(item1 in 1:5){
  beta1[item1] ~ dgamma(4, 3)
  alpha1[item1] ~ dnorm(0, .25)
 }
}
```

Bugs Model Code for Constant Standard Model

```
model{
 for(i in 1:n){# n is the number of obs

# CIRI items, DISAP, KILL, POLPRIS, TORT
   for(item in 1:4){
    logit(Z[i, item, 1])  <- alpha3[item, 1] - beta3[item]*x[i]
    logit(Z[i, item, 2])  <- alpha3[item, 2] - beta3[item]*x[i]
    Pi[i, item, 1] <- Z[i, item, 1]
    Pi[i, item, 2] <- Z[i, item, 2] - Z[i, item, 1]
    Pi[i, item, 3] <- 1 - Z[i, item, 2]
    y[i, item] ~ dcat(Pi[i, item, 1:3])
   }

# PTS Amnesty
    logit(Z[i, 5, 1])  <- alpha5[1, 1] + beta5[1]*x[i]
    logit(Z[i, 5, 2])  <- alpha5[1, 2] + beta5[1]*x[i]
    logit(Z[i, 5, 3])  <- alpha5[1, 3] + beta5[1]*x[i]
    logit(Z[i, 5, 4])  <- alpha5[1, 4] + beta5[1]*x[i]
    Pi[i, 5, 1] <- Z[i, 5, 1]
    Pi[i, 5, 2] <- Z[i, 5, 2] - Z[i, 5, 1]
    Pi[i, 5, 3] <- Z[i, 5, 3] - Z[i, 5, 2]
    Pi[i, 5, 4] <- Z[i, 5, 4] - Z[i, 5, 3]
    Pi[i, 5, 5] <- 1 - Z[i, 5, 4]
    y[i, 5] ~ dcat(Pi[i, 5, 1:5])

# PTS State
    logit(Z[i, 6, 1])  <- alpha5[2, 1] + beta5[2]*x[i]
    logit(Z[i, 6, 2])  <- alpha5[2, 2] + beta5[2]*x[i]
    logit(Z[i, 6, 3])  <- alpha5[2, 3] + beta5[2]*x[i]
    logit(Z[i, 6, 4])  <- alpha5[2, 4] + beta5[2]*x[i]
    Pi[i, 6, 1] <- Z[i, 6, 1]
    Pi[i, 6, 2] <- Z[i, 6, 2] - Z[i, 6, 1]
    Pi[i, 6, 3] <- Z[i, 6, 3] - Z[i, 6, 2]
    Pi[i, 6, 4] <- Z[i, 6, 4] - Z[i, 6, 3]
    Pi[i, 6, 5] <- 1 - Z[i, 6, 4]
    y[i, 6] ~ dcat(Pi[i, 6, 1:5])

# Hathaway
    logit(Z[i, 7, 1])  <- alpha5[3, 1] + beta5[3]*x[i]
    logit(Z[i, 7, 2])  <- alpha5[3, 2] + beta5[3]*x[i]
    logit(Z[i, 7, 3])  <- alpha5[3, 3] + beta5[3]*x[i]
    logit(Z[i, 7, 4])  <- alpha5[3, 4] + beta5[3]*x[i]
    Pi[i, 7, 1] <- Z[i, 7, 1]
    Pi[i, 7, 2] <- Z[i, 7, 2] - Z[i, 7, 1]
    Pi[i, 7, 3] <- Z[i, 7, 3] - Z[i, 7, 2]
    Pi[i, 7, 4] <- Z[i, 7, 4] - Z[i, 7, 3]
    Pi[i, 7, 5] <- 1 - Z[i, 7, 4]
    y[i, 7] ~ dcat(Pi[i, 7, 1:5])

# ITT
    logit(Z[i, 8, 1])  <- alpha6[1, 1] + beta6[1]*x[i]
    logit(Z[i, 8, 2])  <- alpha6[1, 2] + beta6[1]*x[i]
    logit(Z[i, 8, 3])  <- alpha6[1, 3] + beta6[1]*x[i]
    logit(Z[i, 8, 4])  <- alpha6[1, 4] + beta6[1]*x[i]
```

```
    logit(Z[i, 8, 5])  <- alpha6[1, 5] + beta6[1]*x[i]
    Pi[i, 8, 1] <- Z[i, 8, 1]
    Pi[i, 8, 2] <- Z[i, 8, 2] - Z[i, 8, 1]
    Pi[i, 8, 3] <- Z[i, 8, 3] - Z[i, 8, 2]
    Pi[i, 8, 4] <- Z[i, 8, 4] - Z[i, 8, 3]
    Pi[i, 8, 5] <- Z[i, 8, 5] - Z[i, 8, 4]
    Pi[i, 8, 6] <- 1 - Z[i, 8, 5]
    y[i, 8] ~ dcat(Pi[i, 8, 1:6])

# Genocide event data
   logit(p[i,1]) <- alpha1[1] - beta1[1]*x[i]
   y[i, 9] ~ dbern(p[i,1])

# Rummel event data
   logit(p[i,2]) <- alpha1[2] - beta1[2]*x[i]
   y[i, 10] ~ dbern(p[i,2])

# Massive Repression data
   logit(p[i,3]) <- alpha1[3] - beta1[3]*x[i]
   y[i, 11] ~ dbern(p[i,3])

# WHPSI killing  event data
   logit(p[i,4]) <- alpha1[4] - beta1[4]*x[i]
   y[i, 12] ~ dbern(p[i,4])

# UPSALA killing  event data
   logit(p[i,5]) <- alpha1[5] - beta1[5]*x[i]
   y[i, 13] ~ dbern(p[i,5])

# redraw latent variable parameter from mu matrix because of unbalanced panels
   x[i] <- mu[country[i], year[i]]
 }

 sigma ~ dunif(0,1)
 kappa <- pow(sigma, -1)
 for(c in 1:n.country){
  mu[c, 1] ~ dnorm(0, 1)
  for(t in 2:n.year){ #n.year is number of years
   mu[c, t] ~ dnorm(mu[c, t-1], kappa)
  }
 }
 for(item3 in 1:4){
  beta3[item3] ~ dgamma(4, 3)
  alpha03[item3, 1] ~ dnorm(0, .25)
  alpha03[item3, 2] ~ dnorm(0, .25)
  alpha3[item3, 1:2] <- sort(alpha03[item3, 1:2])
 }
 for(item5 in 1:3){
  beta5[item5] ~ dgamma(4, 3)
  alpha05[item5, 1] ~ dnorm(0, .25)
  alpha05[item5, 2] ~ dnorm(0, .25)
  alpha05[item5, 3] ~ dnorm(0, .25)
  alpha05[item5, 4] ~ dnorm(0, .25)
  alpha5[item5, 1:4] <- sort(alpha05[item5, 1:4])
 }
 for(item6 in 1:1){
```

```
  beta6[item6] ~ dgamma(4, 3)
  alpha06[item6, 1] ~ dnorm(0, .25)
  alpha06[item6, 2] ~ dnorm(0, .25)
  alpha06[item6, 3] ~ dnorm(0, .25)
  alpha06[item6, 4] ~ dnorm(0, .25)
  alpha06[item6, 5] ~ dnorm(0, .25)
  alpha6[item6, 1:5] <- sort(alpha06[item6, 1:5])
 }
# Genocide, Rummel, Massive Repression, UCDP killing and WHPSI execution parameters
 for(item1 in 1:5){
  beta1[item1] ~ dgamma(4, 3)
  alpha1[item1] ~ dnorm(0, .25)
 }
}
```

References

Albert, James H. and Val E. Johnson. 1999. *Ordinal Data Modeling*. Springer-Verlag.

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–1288.

Berman, Maureen R. and Roger S. Clark. 1982. "State Terrorism: Disappearances." *Rutgers Law Journal* 13(3):531–578.

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

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.

Carter, David B. and Curt S. Signorino. 2010. "Back to the Future: Modeling Temporal Dependence in Binary Data." *Political Analysis* 18(3):271–292.

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):407–417.

Cingranelli, David L. and David L. Richards. 2012*a*. "The Cingranelli-Richards (CIRI) Human Rights Data Project Coding Manual Version 2008.3.13.".
URL: *http://ciri.binghamton.edu/documentation/ciri_coding_guide.pdf*

Cingranelli, David L. and David L. Richards. 2012*b*. "The Cingranelli-Richards Human Rights Dataset Version 2008.03.12.".
URL: *http://www.humanrightsdata.org*

Clark, Ann Marie. 2001. *Diplomacy of Conscience*. Princeton, NJ: Princeton University Press.

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.

Conrad, Courtenay R. and Will H. Moore. 2011. "The Ill-Treatment & Torture (ITT) Data Project (Beta) Country–Year Data User's Guide." *Ill Treatment and Torture Data Project* .
URL: *http://www.politicalscience.uncc.edu/cconra16/UNCC/Under_the_Hood.html*

Davenport, Christian. 1995. "Multi-Dimensional Threat Perception and State Repression: An Inquiry Into Why States Apply Negative Sanctions." *American Journal of Political Science* 39(3):683–713.

Davenport, Christian. 1997. "From ballots to bullets: an empirical assessment of how national elections influence state uses of political repression." *Electoral Studies* 16(4):517–540.

Davis, David R. and Michael D. Ward. 1990. "They Dance Alone: Deaths and the Disappeared in Contemporary Chile." *Journal of Conflict Resolution* 34(4):449–475.

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

Enterline, Andrew J. and Kristian S. Gleditsch. 2000. "Threats, opportunity, and force: Repression and diversion of domestic pressure, 1948-1982." *International Interaction* 26(1):21–53.

Gastil, Raymond. 1980. *Freedom in the World: Political Rights and Civil Liberties 1980*. New Brunswick, NJ: Transaction Books.

Gelman, Andrew and Donald B. Rubin. 1992. "Inference from iterative simulation using multiple sequences." *Statistical Science* 7:457–511.

Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge, MA: Cambridge University Press.

Gelman, Andrew, John Carlin, Hal Stern and Donald Rubin. 2003. *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC.

Geweke, John. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. Berger, A. P. Dawid and J. F. M. Smith. Oxford: Oxford University Press pp. 169–193.

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

Gibney, Mark and Matthew Dalton. 1996. The Political Terror Scale. In *Human Rights and Developing Countries*, ed. D. L. Cingranelli. Vol. 4 of *Policy Studies and Developing Nations* Greenwich, CT: JAI Press pp. 73–84.

Gibney, Mark and Michael Stohl. 1988. Human Rights and US Refugee Policy. In *Open Borders? Closed Societies? The Ethical and Political Issues*, ed. M. Gibney. New York: Greenwood Press chapter 6, pp. 151–183.

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.

Heidelberger, Philip and Peter D. Welch. 1981. "A spectral method for confidence interval generation and run length control in simulations." *Communications of the ACM* 24:233–245.

Heidelberger, Philip and Peter D. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research* 31(6):pp. 1109–1144.

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

King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.

Lopez, George A. and Michael Stohl. 1992. Problems of Concept and Measurement in the Study of Human Rights. In *Human Rights and Statistics: Getting the Record Straight*, ed. Thomas B. Jabine and Richard P. Claude. Philadelphia, PA: University of Pennsylvania Press.

McCormick, James M. and Neil J. Mitchell. 1997. "Human rights violations, umbrella concepts, and empirical analysis." *World Politics* 49(4):510–525.

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

Mokken, R. J. 1971. *A Theory and Procedure of Scale Analysis*. The Hague: Mouton.

Plummer, Martyn. 2010. "JAGS (Just Another Gibbs Sampler) 1.0.3 Universal.".
URL: *http://www-fis.iarc.fr/ martyn/software/jags/*

Poe, Steven C. 2004. The Decision to Repress: An Integrative Theoretical Approach to the Research on Human Rights and Repression. In *Understanding Human Rights Violations: New Systematic Studies*, ed. S. Carey and S. Poe. Aldershott: Ashgate pp. 16–42.

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., Nicolas Rost and Sabine C. Carey. 2006. "Assessing Risk and Opportunity in Conflict Studies: A Human Rights Analysis." *Journal of Conflict Resolution* 50(4):484–507.

Poe, Steven C. and Rangsima Sirirangsi. 1993. "Human Rights and U.S. Economic Aid to Africa." *International Interactions* 18:309–322.

Ratner, Steve R. and Jason S. Abrams. 2001. *Accountability for Human Rights Atrocities in International Law: Beyond the Nuremberg Legacy*. New York, NY: Oxford University Press.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: J. Wiley & Sons.

Rummel, Rudolph J. 1966. "Dimensions of Conflict Behavior within Nations, 1946-59." *Journal of Conflict Resolution* 10(1):65–73.

Rummel, Rudolph J. 1967. "Understanding Factor Analysis." *Journal of Conflict Resolution* 11(4):444–480.

Rummel, Rudolph J. 1994. "Power, Geocide and Mass Murder." *Journal of Peace Research* 31(1):1–10.

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

Rummel, Rudolph J. and Raymond Tanter. 1974. "Dimensions of Conflict Behavior Within and Between Nations, 1955-1960." *ICPSR 5021* .

Schnakenberg, Keith E. and Christopher J. Fariss. 2013. "Dynamic Patterns of Human Rights Practices." *Political Science Research and Methods* DOI: http://dx.doi.org/10.1017/psrm.2013.15.

Sinkovits, Robert S., Pietro Cicotti, Shawn Strande, Mahidhar Tatineni, Paul Rodriguez, Nicole Wolter and Natasha Bala. 2011. "Data intensive analysis on the gordon high performance data and compute system." *KDD '11 Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* pp. 747–748.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin and Angelika Van Der Linde. 2002. "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4):583–639.

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.

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.