UNIVERSITY OF CALIFORNIA, SAN DIEGO

The More Things Stay the Same the More They Change: Measuring Changing Levels of Human Rights Using Computational Methods

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Political Science

by

Christopher J. Fariss

Committee in charge:

Professor James H. Fowler, Chair
Professor David A. Lake, Co-Chair
Professor Jesse Driscoll
Professor Miles Kahler
Professor Brice X. Semmens

2013

The dissertation of Christopher J. Fariss is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2013

DEDICATION

To my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am deeply indebted to a large number of individuals who supported me while I wrote each of the chapters in my dissertation. I am particularly thankful for the advice and comments that I received from the faculty and students of the Political Science Department at the University of California, San Diego and at the University of North Texas. Special thanks goes to my dissertation advisors at UCSD, James Fowler, David Lake, Miles Kahler, Jesse Driscoll, and Brice Semmens. Each of these individuals have been generous with their time and advice.

I would also like to thank my thesis advisors at UNT, Jim Meernik, Steve Poe, Kimi King and David Mason. Steve Poe first introduced me to the measurement of human rights when I was an undergraduate at the University of North Texas in the summer of 2003. That class sparked an interest that has been with me for the last decade. Steve passed away in August 2007.

Special thanks also goes to my friend and collaborator Keith Schnakenberg. We began working together during my first week as a graduate student at UCSD in James Fowler's social network analysis class. Keith was taking classes at UCSD while finishing his master's thesis at San Diego State University. Our first collaboration turned into chapter 2 of this dissertation and is forthcoming in the *Journal of Conflict Resolution*. Over the last five years, we have collaborated on several additional projects and plan many more.

Bob Sinkovits, Paul Rodriguez, and Nicole Wolter helped me get up and running on the Gordon Supercomputer, which is located at the San Diego Supercomputer Center at UCSD. Without their help, the estimation of some of the models discussed in my dissertation would have taken considerably longer to finish.

There are many other colleagues that I wish to thank. I hope that I have not missed anyone. Specifically, I would like to thank Clair Apodaca, Megan Becker, Sam Bell, Mathew Bergman, Rob Bond, Nicole Bonoff, Cameron Brown, Craig Burnett, Francisco Cantu, Rumman Chowdhury, Chad Clay, Courtenay Conrad, Cesi Cruz, Saul Cunow, Geoff Dancy, Christian Davenport, Michael Davidson, Chris Dawes, Elaine Denny, Scott Desposato, Dan Enemark, Andrew Enterline, Matthew Eshbaugh-Soha, Erik Gartzke, Laura Gee, Micah Gell-Redman, Peter Gourevitch, Benjamin Graham, Michael Greig, Dimitar Gueorguiev, Scott Guenther, Hans Hassell, Cullen Hendrix, Seth Hill, Will Hobbs, Jeremy Horowitz, Alex Hughes, Ryan Jablonski, Gary Jacobson, Jason Jones, Jeff Kaplow, Linda Camp Keith, Sarah Knoesen, Vlad Kogan, Matthew Krain, Bethany Lacina, Don Lee, John LeJeune, Brad LeVeck, Yuhui Li, Arend Lijphart, Adeline Lo, Lydia Lundgren, Yon Lupu, Barbara Lutz, Ko Maeda, Dan Maliniak, Scott Martin, Blake McMahon, Jonathan Markowitz, Kelly Matush, Jamie Mayerfeld, Rupal Mehta, Will Moore, Amanda Murdie, Lindsay Nielson, Chris O'Keefe, Nicholas Obradovich, Liz Oldmixon, Steven Oliver, Maya Oren, Kai Ostwald, Kristen Parks, Michael Plouffe, Keith Poole, Justin Reeves, Mike Rivera, Patrick Rogers, Andy Rottman, Alex Ruiz, Sebastian Saiegh, Idean Salehyan, Christina Schneider, Darren Schreiber, Paul Schuler, Deborah Seligsohn, Jaime Settle, Branislav Slantchev, Devesh Tiwari, Mona Vakilifathi, Neil Visalvanich, Andrew Waugh, Wendy Wong, Reed Wood, Thorin Wright, Grace Wu, and

Langche Zeng for many helpful comments and suggestions. I wish there was room to acknowledge each of you more thoroughly.

Finally, I would not have been able to write this dissertation without the support of my parents, Ron and Peggy, my brother and sister-in-law, Ryan and Angie, my friend, Yumi Ogawa, and the rest of my friends and family. To each and every one of you, thank you again for your support.

VITA

2013	Ph.D. in Political Science, University of California-San Diego
2007	M.S. in Political Science, University of North Texas
2005	B.A., in Political Science, University of North Texas
2005	B.F.A. in Drawing and Painting, University of North Texas

PUBLICATIONS

“Dynamic Patterns of Human Rights Practices”, with Keith E. Schnakenberg.
Political Science Research and Methods (Forthcoming)

“Measuring Mutual Dependence Between State Repressive Actions”, with Keith E. Schnakenberg.
Journal of Conflict Resolution (Forthcoming).

“Going the Distance: The Price of Projecting Power”, with Jonathan N. Markowitz.
International Interactions 39(2): 119-143 (April 2013).

“Yahtzee: An Anonymized Group Level Matching Procedure”, with Jason J. Jones, Robert M. Bond, Jaime E. Settle, Adam D. I. Kramer, Cameron Marlow, and James H. Fowler.
PLoS ONE 8(2): e55760 (February 2013).

“Inferring Tie Strength from Online Directed Behavior”, with Jason J. Jones, Jaime E. Settle, Robert M. Bond, Cameron Marlow, and James H. Fowler.
PLoS ONE 8(1): e52168 (January 2013).

“A 61-Million-Person Experiment in Social Influence and Political Mobilization”, with Robert M. Bond, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler.
Nature 489(7415): 295-298 (13 September 2012).

“Emergency and Escape: Explaining Why States Derogate from Human Rights Treaties during National Emergencies”, with Emilie Hafner-Burton and Laurence R. Helfer.
International Organization 65(4): 673-707 (October 2011).

“The Strategic Substitution of United States Foreign Aid"
Foreign Policy Analysis 6(2): 107-131 (April 2010).

“The Path to Asylum in the US and the Determinants for Who gets in and Why", with Andy J. Rottman, and Steven C. Poe.
International Migration Review 43(1): 3-33 (March 2009).

“International Criminal Defense: A Preliminary Analysis of Background and Resources on Judicial Decision Making at the ICTR", with James D. Meernik.
Judicature 89(6): 326-333 (May-June 2006).

ABSTRACT OF THE DISSERTATION

The More Things Stay the Same the More They Change: Measuring Changing Levels of Human Rights Using Computational Methods

by

Christopher J. Fariss

Doctor of Philosophy in Political Science

University of California, San Diego, 2013

Professor James H. Fowler, Chair
Professor David A. Lake, Co-Chair

According to indicators of political repression currently used by scholars, human rights practices have not improved over the past 35 years, despite the spread of human rights norms, better monitoring, and the increasing prevalence of electoral democracy. I argue that this empirical pattern is not an indication of stagnating human rights practices. Instead, it reflects a systematic change in the way monitors encounter and interpret information about abuses. The standard of accountability used to assess state behaviors becomes more stringent as monitors look harder for abuse, look in more places for abuse, and classify more acts as abuse. In chapter 1, I present a new, theoretically informed measurement model, which generates unbiased estimates of repression. I also show that respect for human rights has improved over time and that the relationship between human rights respect and ratification of the UN Convention Against Torture is positive, which contradicts findings from existing research. In chapter 2, I demonstrate other modeling techniques for measuring human rights. In chapter 3, I demonstrate that the ratification of human rights

treaties is empirically associated with higher levels of respect for human rights over time and across countries. This positive relationship is robust to a variety of measurement strategies and model specifications. Overall, a new picture emerges of improving levels of respect for human rights, which coincides with the increasing embeddedness of countries within the international human rights regime. In chapter 4, I extend the model and estimate the distribution of the number of individuals killed for each country-year observation in one of the original event-based datasets. The model explicitly accounts for the uncertainty inherent in counting this type of difficult to observe event. To validate the new model, I focus on one dataset, which defines one-sided government killing as government caused deaths of non-combatants. Up to date versions of each chapter in this dissertation will be made publicly available at my SSRN page: `http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=1320749`. The estimates from each chapter along with the code necessary to implement the models in JAGS and R will be made publicly available at my Dataverse page: `http://dvn.iq.harvard.edu/dvn/dv/HumanRightsScores`.

Chapter 1

Respect for Human Rights has Improved Over Time: A Dynamic Latent Variable Model

1.1 Introduction

Have levels of political repression changed? "Repression" or violations of "physical integrity rights" include arrests and political imprisonment, beatings and torture, extrajudicial executions, mass killing and disappearances, all of which are practices used by political authorities against those under their jurisdiction.[1] This question is important because current indicators of political repression imply that human rights practices have been essentially constant over the last 35 years (see Figure 1.1), despite the spread of human rights norms, better monitoring by private and public agencies, and the increasing prevalence of electoral democracy. While some theorists take issue with this empirical pattern and the data used to support it[2], hundreds of studies rely on these indicators to analyze the determinants of repression[3] and the effects of international institutions on human rights treaty compliance[4].

I argue that this pattern of constant abuse is not an indication of stagnating human rights practices. Instead, it reflects a systematic change in the way monitoring agencies, like Amnesty International and the US State Department, encounter and interpret information about human rights abuses. Over time, this process has led to what I call a *changing standard of accountability*. As a consequence of this change, human rights reports have become increasingly stringent assessments of state behaviors. This change occurs because of (1) the incentive to hide the use of these policy tools by government authorities and (2) the countervailing strategies used by observers and activists interested in revealing, understanding and ultimately changing repressive practices for the better. This interaction between state actors and observers, both academic and activist, affects the production of information used by researchers to quantify repressive behaviors.[5]

I present results for a new view of repression: *physical integrity practices have improved over time*. To support my claim, I compare an existing dynamic ordinal item response theory model, which I call the ***constant standard model***, to a new extension of this model, which I call the ***dynamic standard model***. The new model formalizes the relationship between the unmeasured standard of accountability and observed levels of repression measured by several existing data sources. Note that both of these models are dynamic with respect to the estimated country-year latent variable. The models differ with respect to the standard of accountability. The constant standard model, like all the existing models of repression (e.g., the CIRI Additive index and the Political Terror Scale index), implicitly assumes that the standard of accountability does not change over time. By comparing the constant standard model, which makes this assumption, and the dynamic standard model, which relaxes it, I am able to demonstrate that

[1] This definition is a modified version of one from Goldstein (1978). His definition includes censorship which I exclude in order to focus exclusively on physical integrity violations, which are the most commonly analyzed rights.

[2] See the discussion in Clark and Sikkink (Forthcoming), and Goodman and Jinks (2003).

[3] See for example the research by Bell, Clay and Murdie (2012), Bueno De Mesquita et al. (2005), Cingranelli and Filippov (2010), Conrad and Moore (2010), Davenport (1995), Davenport (2010), Davenport and Armstrong (2004), Fariss and Schnakenberg (2013), Poe and Tate (1994), Poe, Tate and Keith (1999), Wood (2008), and Zanger (2000). Also see the reviews by Davenport (2007*a*) and Poe (2004).

[4] See for example the research by Hathaway (2002), Hafner-Burton and Tsutsui (2005), Hill Jr. (2010), Keith (1999), Keith, Tate and Poe (2009), Lupu (2013*a*), Lupu (2013*b*), Neumayer (2005), and Simmons (2009).

[5] Though human rights theorists are aware of this issue (Brysk 1994; Clark 2001; Goodman and Jinks 2003; Keck and Sikkink 1998), this is the first project that systematically incorporates it into a measurement model of repression.

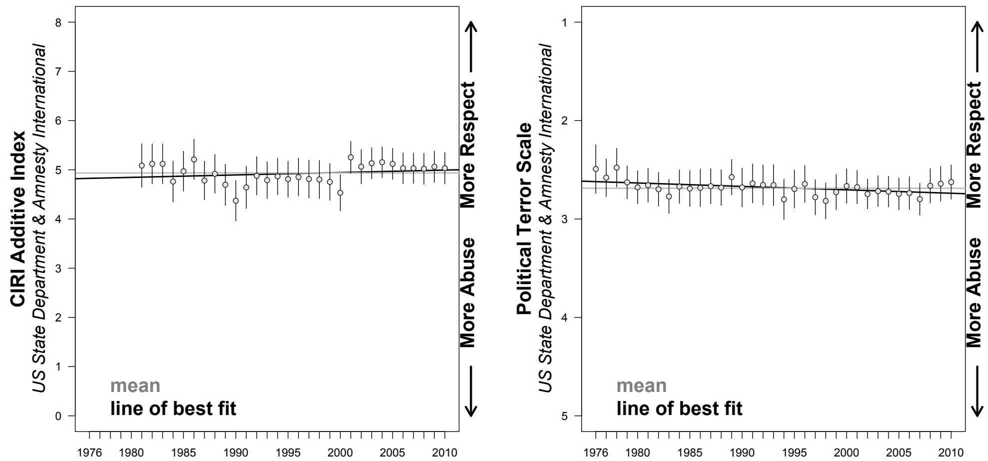


Figure 1.1: Yearly mean and 95% confidence intervals for the estimated level of repression using the CIRI Additive index (left), and the Political Terror Scale index (right). Each series is based on the human rights reports from the US State Department and Amnesty International. Note that the averages for the Political Terror Scale estimates are based on two scales coded independently, one from the US State Department reports and one from the Amnesty International reports. Similar figures for the individual PTS variables are displayed in Section 10.4 of the appendix.

the standard of accountability has become a more stringent assessment of state behaviors over time and that these unaccounted for changes explain why average levels of repression appear stagnant.

I make several important contributions in this paper: First, I develop a theory of the standard of accountability and a new measurement model that accounts for it. Second, the measurement model itself is the first in the political science literature to estimate time-varying item-difficulty cut-points for some items (repression outcome variables). These model parameters measure changes to the standard of accountability over time. Third, I introduce new unbiased estimates of repression that cover the time period beginning in 1949 and ending in 2010 (n= 9267). The resulting data are the most comprehensive estimates of repression that currently exist. Fourth, I provide empirical evidence that human rights practices have improved over time. Fifth, I illustrate the substantive importance of the results to international relations theory by showing that the relationship between human rights respect and ratification of the UN Convention Against Torture is positive, which contradicts findings from existing research. Finally, I demonstrate how to correct for temporal bias in standard models of repression that can accommodate the original ordered variables derived from human rights reports.

1.2 Why the Standard of Accountability Changes Over Time

The standard of accountability, which I define as the set of expectations that monitoring agencies use to hold states responsible for repressive actions, has not been systematically addressed because much of the data measuring repression are derived from the same primary sources (Cingranelli and Richards

1999, 2012*a*,*b*; Gibney, Cornett and Wood 2012; Hathaway 2002). The documents used to measure repression are *The Country Reports on Human Rights Practices* published annually by the US State Department and *The State of the World's Human Rights* report published annually by Amnesty International. The information captured in these documents will bias assessments of repression over time if changes in the standard of accountability are not also taken into account. In the language of research design, *instrumentation bias* occurs if the measurement tool used to assess a behavior changes over time (Trochim and Donnelly 2008).

I argue that the standard of accountability has changed due to a combination of three mechanisms. These mechanisms influence the strategies and therefore the set of expectations that monitoring agencies use to assess and document state behaviors. First, improvements in the quality and increases in the quantity of information have led to more accurate assessments of the conditions in each country over time. Second, access to countries by NGOs, like Amnesty International and Human Rights First (formerly the Lawyer's Committee for Human Rights), which seek to collect and disseminate accurate information about repression allegations and practices has increased as these organizations grow and cooperate with one another. Third, changes in the subjective views of what constitute a "good" human rights record held by analysts at the monitoring agencies are anchored by the status quo, which improves as the global average of rights respect improves.

Human rights theorists recognize that the information used to assess government behaviors may change over time and that this could mask underlying improvements in human rights practices.[6] Keck and Sikkink (1998) attribute this change to an "information paradox". The paradox occurs when an increase in information leads to difficulties in assessing the efficacy of advocacy campaigns over time because of the very success in collecting and aggregating accounts of repressive actions in the first place. Clark and Sikkink (Forthcoming) coin a similar term — "human rights information paradox" — to describe this issue as it relates to human rights abuses specifically. As a result of this paradox, the global human rights situation may appear to have worsened over time because there is simply an increasing amount of information with which to assess human rights practices (see Section 10.3 of the appendix for examples).

Innes de Neufville (1986) argues that the quality of the human rights reports produced by the US State Department increased because of changes to the reporting requirements, which "altered practices and norms within the Department of State and created an arena for public evaluation of the information" (682). The improvement in the quality of these reports is corroborated by yearly critiques published by the Lawyers Committee for Human Rights. This improvement has also been documented by an analysis of an index derived from the State Department reports compared to the same index based on reports from Amnesty International (Poe, Carey and Vazquez 2001).

In addition to the quality and quantity of information, access to government documentation, witnesses, victims, prisons sites, and other areas are important for assessing state behaviors. Both Amnesty International and the US State Department rely on reports from other NGOs that collect and disseminate

[6] See for example the research by Bollen (1986), Brysk (1994), Clark (2001), Goodman and Jinks (2003), and Keck and Sikkink (1998).

information about human rights abuses within states. The number and effectiveness of these actors has increased over time, especially since the end of the Cold War.[7] Moreover, as Hill Jr., Moore and Mukherjee (Forthcoming) argue, increasing numbers of domestic NGOs generate more credible signals about government abuse which are used by Amnesty International and by extension the US State Department in the production of human rights reports.

Monitoring agencies are also increasingly sensitive to the various kinds of ill-treatment that previously fell short of abuse but that still constitute violations of human rights. There is specific evidence from case law of a rising standard of acceptable treatment, whereby more acts come to be classified as inhuman treatment or torture. For example the European Court of Human Rights, in Selmouni v. France (1999), "consider certain acts which were classified in the past as "inhuman and degrading treatment" as opposed to "torture" could be classified differently in future." That is, acts by state agents that might have previously been classified within the less severe category of ill-treatment and degrading punishment might now be classified as torture. The court states further "that the increasingly high standard being required in the area of the protection of human rights and fundamental liberties correspondingly and inevitably requires greater firmness in assessing breaches of the fundamental values of democratic societies."[8]

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. For example, Amnesty International expanded its strategy over time as it responded to developments in the repressive behaviors used by states.[9] The initial focus of Amnesty International on political prisoners during the 1960s and 1970s precluded the reporting of extrajudicial killings that took place outside of prisons (Clark, 2001, ch. 5). Also during the 1960s and 1970s, state agents in Guatemala frequently disappeared opposition members, yet Amnesty International did not document these policies until 1976 because these actions were not initially a policy tool of concern (Clark, 2001, ch. 4).

Thus, the set of expectations that monitoring agencies use to hold states responsible for repressive actions changes over time. The reports published today represent a broader and more detailed view of the human rights practices than reports published in previous years. As Sikkink notes, these monitoring agencies and others "have expanded their focus over time from a narrow concentration on direct government responsibility for the death, disappearance, and imprisonment of political opponents to a wider range of rights, including the right of people to be free from police brutality and the excessive use of lethal force" (2011, 159).

Unfortunately for scholars interested in these changes, the standard of accountability is not directly observable in human rights reports and is therefore difficult to measure. To make matters more complicated, alternative sources of information that were once highly cited are now largely forgotten and

[7]See discussions in Hopgood (2006), Hill Jr., Moore and Mukherjee (Forthcoming), Korey (2001), Keck and Sikkink (1998), Lake and Wong (2009), Murdie and Bhasin (2011), Murdie and Davis (2012), and Wong (2012).

[8]Selmouni v. France, 25803/94, Council of Europe: European Court of Human Rights, 28 July 1999, available at: `http://www.unhcr.org/refworld/docid/3ae6b70210.html`.

[9]See Clark (2001) for a discussion of the developments in the strategy used by Amnesty International in response to changes in repressive behaviors. Berman and Clark (1982) provides an example of how political authorities in the Philippines began to disappear political opponents to avoid public scrutiny of other human rights violations.

out of date (Harff and Gurr 1988; Rummel 1994*a*,*b*, 1995; Taylor and Jodice 1983). Contemporary alternatives often cover shorter periods of time (Conrad, Haglund and Moore 2012; Eck and Hultman 2007), are not up to date (Conrad, Haglund and Moore 2012; Hathaway 2002) or still rely on the same standards-based human rights reports (Hathaway 2002). All of these issues make the systematic comparison of results from different data sources difficult, which leaves the problem of instrumentation bias acknowledged but unaddressed in the literature. None of the issues are cause for concern however, because the computational tools necessary to link diverse sources of data with theory now exist. The latent variable models I describe below are capable of (1) bringing together diverse sources of information, (2) assessing the relative quality of the information included, and (3) quantifying the certainty of the estimates of repression that are generated from the models. These models allow me to test for the influence of the standard of accountability by comparing the new model in which the probability of documenting a repressive action changes over time (dynamic standard model) to the existing model in which this probability does not change (constant standard model). In the next section I introduce and make a theoretical distinction between standards-based data and event-based data. I then introduce the latent variable models.

1.3 Standards-Based Data and Event-Based Data

Before discussing the theoretical differences between standards-based data and event-based data the reader should keep in mind that all of the variables included in the two competing latent variable models are same and are operationalized to capture one or more of the repressive behaviors identified in the definition of repression used throughout this paper. Recall that the definition of "repression" or violations of "physical integrity rights" and sometimes called "state sanctioned terror" includes arrests and political imprisonment, beatings and torture, extrajudicial executions, mass killing and disappearances, all of which are practices used by political authorities against those under their jurisdiction. The different repressive tactics are related to one another in that they are all used to mitigate potential threats to the survival of the regime. To address such threats, some leaders choose the complete elimination of a political group (politicide) or other group designation (genocide) (Harff 2003; Rummel 1994*b*, 1995). Massive repressive events are related to the intent of genocide in the use of mass slaughters or pogroms to eliminate substantial portions of a predetermined group but is a broader category that includes a greater number of events than genocide or politicide (Harff and Gurr 1988). The use of extrajudicial killing to eliminate individuals is captured by both the political execution data (Taylor and Jodice 1983) and several of the variables derived from the human rights reports (Cingranelli and Richards 2012*a*,*b*; Gibney, Cornett and Wood 2012). The measurement of one sided government killing in which more than 25 individuals (non-combatants) are killed excludes extrajudicial killings that occur inside a prison and combatant deaths that occur during civil conflicts (Eck and Hultman 2007). Each of these tactics can be used to target individuals or groups that a regime views as a threat.

Standards-based variables, capture individually or as part of an index, extrajudicial killing, torture, political imprisonment, and disappearances (Cingranelli and Richards 2012*a*,*b*; Conrad and Moore

2011; Conrad, Haglund and Moore 2012; Gibney, Cornett and Wood 2012; Hathaway 2002). 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. The standards-based and event-based variable names, operationalizations, citations and data sources are displayed in Table 4.5 and Table 4.6 respectively. In sum, all of these variables capture the use of "repression" or violations of "physical integrity rights". The temporal coverage and data type of each variable are displayed in Figure 1.2.

If the standard of accountability has increased over time, then comparisons of data derived from standards-based documents will be biased over time because of unaccounted for changes in the instrument (human rights documents) used to measure behavioral change (levels of repression). Temporal comparisons of categorical measures derived from the standards-based data sources are problematic because the data are based on content of reports that were prepared in a specific historical context. The reports are primary source documents that are used by analysts to derive variables on repression. The issue of temporal comparability arises because the older reports are not updated or revised even if new information about specific repressive events is obtained over time or as the goals, strategic incentives, or status quo expectations of the monitoring agencies evolve. These same issues make data derived from these reports quite useful for comparing state behaviors in the same year.

Event-based data sources contrast with the standards-based data because they are based on evidence derived from a pool of regularly augmented primary source documents. However, it is also likely that increased access to countries over time will also lead to an increase in the accuracy of the event-count data. The producers of the event-based data are aware of this process. When new information about repressive actions becomes available from NGOs, news reports, historians, or truth commissions, these scholars update their data. Moreover, information from multiple sources are used to help corroborate each datum.These data therefore represent the best approximation of the historical pattern of repression for a given country at each update. For example, Rummel discusses the process by which he periodically updated the event-based information used in his articles and books.[10] Similar discussions can be found in the documentation of the other event-based data.

Skepticism over the comparability of event data that counts the number of repressive events in country-year observations was one of the main reasons for the movement away from event data in cross-national human rights research.[11] 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-

[10] See the discussion in the preface of Rummel's book *Death by Government: Genocide and Mass Murder in the Twentieth Century* (1994*a*, xi-xxii) as well as his other books about specific cases. Much of this material is publicly available at Rummel's website: `http://www.hawaii.edu/powerkills/welcome.html`.

[11] See Poe (2004) for a review of the literature critiquing event-based data. Brysk (1994) provides a specific example critical of comparisons of event-based data.

based data.[12] I avoid this issue for now by focusing on event data that are binary.[13] However, this raises another practical issue: binary event data only capture extreme levels of abuse. These data have been useful for comparing broad trends but the relative specificity of the standards-based reports is another reason for their preeminence over the event-based binary data. The standards-based data have provided analysts with more behavioral categories for comparison. The models I develop below can incorporate information from multiple sources, and quantify the uncertainty of each estimate, conditional on the availability of each variable included in the model. Missing data does not lead to a loss of country-year-observations but only increases the uncertainty for the estimate of a given country-year. The models can also accommodate variables measured using different scales. Section 10.1 and 10.2 of the appendix contains additional information about the development and coding rules for these variables.

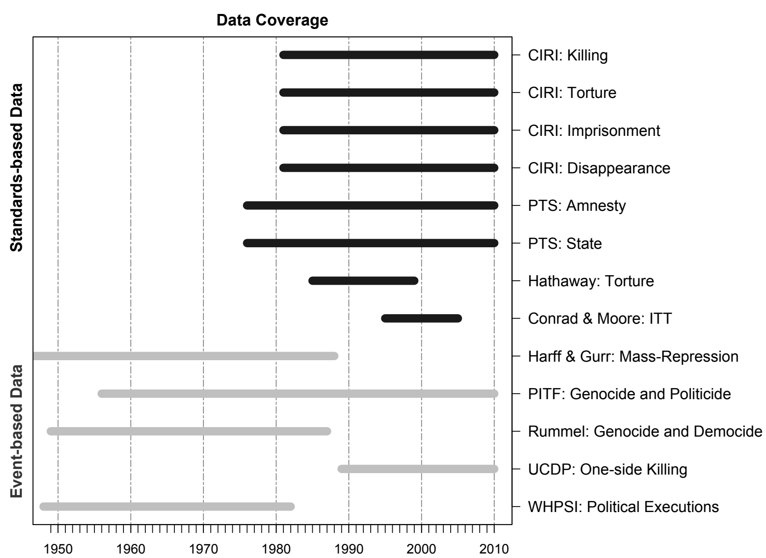


Figure 1.2: Temporal coverage and data type of repression data sources. See Table 4.5 and Table 4.6 for more information. Grey lines are event-based data. Black lines are standards-based measures.

[12]The seminal work of Lars Schoultz (1981) was the first quantitative test between the stated importance of human rights by the United States government and the allocation of foreign aid using event-based human rights data for Latin American states. The use of event-based counts was criticized (e.g., Carleton and Stohl 1985; Stohl, Carleton and Johnson 1984) and led to an debate about the pros and cons of event-based and standards-based variables. 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).

[13]The models I describe below can be extended to incorporate the actual event counts. This project is currently underway.

Table 1.1: Standards-Based Repression Data Sources

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
CIRI Physical Integrity Data, 1981-2010 - political imprisonment (ordered scale, 0-2) - torture (ordered scale, 0-2) - extrajudicial killing (ordered scale, 0-2) - disappearance (ordered scale, 0-2)	Cingranelli and Richards (1999, 2012*a*,*b*) Amnesty International Reports[1] and State Department Reports[2] *Information in Amnesty reports takes precedence over information in State Department reports*
Hathaway Torture Data, 1985-1999 - torture (ordered scale, 1-5)	Hathaway (2002) State Department Reports[1]
Ill-Treatment and Torture (ITT), 1995-2005 - torture (ordered scale, 0-5)	Conrad and Moore (2011), Conrad, Haglund and Moore (2012), Amnesty International (2006) Annual Reports[1], press releases[1], and Urgent Action Alerts[1]
PTS Political Terror Scale, 1976-2010 - Ammesty International scale (ordered scale, 1-5) - State Department scale (ordered scale, 1-5)	Gibney, Cornett and Wood (2012), Gibney and Dalton (1996) Amnesty International Reports[1] State Department Reports[1]

1. Primary Source; 2. Secondary Source

Table 1.2: Event-Based Repression Data Sources

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
Harff and Gurr Dataset, 1946-1988 - massive repressive events (1 if country-year experienced event 0 otherwise)	Harff and Gurr (1988) historical sources (see article references)[1]
Political Instability Task Force (PITF), 1956-2010 - genocide and politicide (1 if country-year experienced event 0 otherwise)	Harff (2003), Marshall, Gurr and Harff (2009) historical sources (see article references)[1] State Department Reports[2] Amnesty International Reports[2]
Rummel Dataset, 1949-1987 - genocide and democide (1 if country-year experienced event 0 otherwise)	Rummel (1994*b*, 1995), Wayman and Tago (2010) New York Times[1], New International Yearbook[2], Facts on File[2], Britannica Book of the Year[2], Deadline Data on World Affairs[2], Kessing's Contemporary Archives[2]
UCDP One-sided Violence Dataset, 1989-2010 - government killing (event count estimate) (1 if country-year experienced event 0 otherwise)	Eck and Hultman (2007), Sundberg (2009) Reuters News[1], BBC World Monitoring[1] Agence France Presse[1], Xinhua News Agency[1], Dow Jones International News[1], UN Reports[2], Amnesty International Reports[2], Human Rights Watch Reports[2], local level NGO reports (not listed)[2]
World Handbook of Political and Social Indicators WHPSI, 1948-1982 - political executions (event count estimate) (1 if country-year experienced event 0 otherwise)	Taylor and Jodice (1983) New York Times[1], Middle East Journal[2], Asian Recorder[2], Archiv der Genenwart[2] African Diary[2], Current Digest of Soviet Press[2]

1. Primary Source; 2. Secondary Source

1.4 Two Competing Latent Variables Models of Repression

The latent variable models I develop in this paper are item-response theory (IRT) models. The dynamic standard model is an extension of the DO-IRT (dynamic ordinal item response theory) model developed by Schnakenberg and Fariss (2012). The constant standard model is identical to the DO-IRT model. Note that both models presented in this paper are dynamic with respect to the estimated latent human rights variable. The models differ with respect to the standard of accountability. In one case the standard changes (dynamic standard model) and in one case it does not change (constant standard model).

The constant standard model in addition to all of the existing models of repression — those based on information from the annual human rights reports — implicitly assume a constant standard of accountability over time. By comparing the estimates from the constant standard model, which makes this assumption, and the dynamic standard model, which relaxes this assumption, I am able to test the hypothesis that an increase in the standard of accountability — the probability of observing and therefore coding a repressive outcome — increases over time for the repression variables derived from the human rights reports.

Schnakenberg and Fariss (2012) model the latent respect for human rights for a country in a particular year as dependent on the value for the same country in the previous year. These authors demonstrate that the dynamic latent variable model fits the CIRI human rights data (Cingranelli and Richards 2012*a*,*b*) substantially better than a static latent variable model similar to those used in the democracy literature.[14] The latent variable models measuring democracy developed by Treier and Jackman (2008) and Pemstein, Meserve and Melton (2010) assume that the observed indicators used in the model are independent conditional on the value of the trait to be estimated, which is an overly strong assumption in the case of human rights as demonstrated by Schnakenberg and Fariss (2012). And, though Armstrong (2011) relaxes this assumption by using a dynamic factor analytic model to analyze the Freedom House Indicators, he models the observed indicators as interval response variables instead of ordered categories and provides no evidence that the model performs better than any alternative parameterizations.[15] As I demonstrate below, model comparison statistics represent the best way to adjudicate between competing theories and the measurement models deduced from them.

To parameterize the changing standard of accountability, I allow the baseline probability of observing a given level of repression for a specific repression variable or *item* to vary as a function of time in one model (dynamic standard model) and compare the resulting estimates to another in which this probability is constant (constant standard model).[16] This is accomplished by estimating time varying "item difficulty cut-points" or "thresholds" for some of the items. These parameters are analogous to the

[14] Note that both of the models compared by Schnakenberg and Fariss (2012) assume a constant standard of accountability.

[15] All of this research builds on the seminal work by Poole and Rosenthal (1991, 1997), which employs a maximum likelihood approach to model political ideology from roll call votes. See Clinton, Jackman and Rivers (2004) and Martin and Quinn (2002) for Bayesian implementations of this model. See Jackman (2008) for a thorough discussion of the development of these and other measurement models.

[16] I use the term "item" and "variable" interchangeably throughout this paper. The term "item" is attributed to researchers developing educational tests (e.g., Lord 1980; Lord and Novick 1968; Rasch 1980). See Borsboom (2005) and Jackman (2008) for accounts of the development of this literature.

intercept term in a linear model. The rest of the model parameters are similar to other latent variable models in the literature and are described in detail below. Thus, the changing standard or accountability is parameterized in the dynamic standard model by estimating the item difficulty cut-points for the data sources that are derived from information contained in the annual human rights reports. Constant item difficulty cut-points are estimated for the event-based data sources. This parameterization is motivated by the theoretical distinction between the standards-based data sources and event-based data sources. Again, both models are dynamic with respect to the latent variable itself.

Formally, the statistical models I compare in this paper are both built on the assumption that the observed repression outcome variables for the country-year observations are each a function of the same underlying unidimensional latent variable, which represents the "true" or "latent" level of repression or respect for physical integrity rights. The goal of these models is to estimate θ_{it}, which is the latent level of respect for physical integrity rights of country i in year t. For each model there are J indicators $j = 1,\ldots,J$. Some of the j indicators are ordinal with varying number of levels and some of the j indicators are binary. As already noted, $i = 1,\ldots,N$ indexes cross-sectional units and $t = 1,\ldots,T$ indexes time periods. y_{itj} is observed for each of the $j = 1,\ldots,J$ indicators displayed in Table 4.5 and Table 4.6. Each indicator is either ordinal or binary and can take on K_j values. For the binary indicators, $K_j = 2$.

For each item, there is an "item discrimination" parameter β_j and a set of $K_j - 1$ "item difficulty cut-points" $\left(\alpha_{jk}\right)_{k=1}^{K_j}$. These parameters are analogous to a slope and intercept term in a logistic regression or the slope and cut-points in an ordered logistic regression.

For the dynamic standard model, I specify the parameterization of the difficulty cut-points for some of the items to vary over time such that $\left(\alpha_{tjk}\right)_{k=1}^{K_j}$. Note the t subscript here. This parameterization includes the standards-based variables from Cingranelli and Richards (1999, 2012*a*,*b*), Gibney, Cornett and Wood (2012), Hathaway (2002), and Conrad, Haglund and Moore (2012). The other items retain the constant item difficulty cut-point parameterization: $\left(\alpha_{jk}\right)_{k=1}^{K_j}$, which include the binary event-based variables drawn from Harff and Gurr (1988), Harff (2003), Rummel (1994*b*, 1995), Eck and Hultman (2007), Taylor and Jodice (1983).[17] Note the lack of a t subscript here. There is no t subscript on this parameter for any of the items in the constant standard model.

I assume error terms ε_{itj} are independently drawn from a logistic distribution, where $F(\cdot)$ denotes the logistic cumulative distribution function.The probability distribution for a given response to item j in the constant standard model is therefore given by

$$P[y_{itj} = 1] = F(\alpha_{j1} - \theta_{it}\beta_j) \quad (1.1)$$

$$P[y_{itj} = k] = F(\alpha_{jk} - \theta_{it}\beta_j) - F(\alpha_{jk-1} - \theta_{it}\beta_j) \quad (1.2)$$

$$P[y_{itj} = K] = 1 - F(\alpha_{jK-1} - \theta_{it}\beta_j) \quad (1.3)$$

[17] It is a coincidence that the event-based variables are each binary whereas the standards-based data are all categorical. The model is not dependent on this distinction.

For each item with constant difficulty cut-points, $y_{itj} = k$ if $\alpha_{jk-1} < \theta_{it}\beta_j + \varepsilon_{itj} < \alpha_{jk}$, and by specifying $\alpha_{j0} = -\infty$ and $\alpha_{jK_j} = \infty$ the probability equations (1), (2), and (3) reduce to[18]

$$P[y_{itj} = k] = F(\alpha_{jk} - \theta_{it}\beta_j) - F(\alpha_{jk-1} - \theta_{it}\beta_j) \tag{1.4}$$

Therefore, assuming local independence of responses across units, the constant standard's likelihood function for β, α, and θ given the data is $\mathscr{L}(\beta, \alpha, \theta | y)$ and is expressed as

$$\mathscr{L} = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J} \left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j)\right] \tag{1.5}$$

The first set of equations (1), (2), and (3) and the reduced form (4) refer to the probability of observing a particular hypothetical level k. The likelihood equation (5) refers to the probability of the observed level in the data y_{itj}. These equations are the same for the dynamic standard model except for the addition of the t subscript on some of the α parameters. As a notational convenience let $v_j = 1$ when the j indicator is one of the standards-based variables and then $v_j = 0$ when it is one of the event-based variables. The probability distribution for the dynamic standard model is therefore

$$P[y_{itj} = k] = \left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j)\right]^{(v_j)} * \left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j)\right]^{(1-v_j)} \tag{1.6}$$

And the dynamic standard's likelihood function $\mathscr{L}(\beta, \alpha, \theta | y)$ is expressed as

$$\mathscr{L} = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J} \left[F(\alpha_{tjy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{tjy_{itj}-1} - \theta_{it}\beta_j)\right]^{(v_j)} * \left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j)\right]^{(1-v_j)} \tag{1.7}$$

Note that when $v_j = 0$, the probability distribution (6) and the likelihood function (7) for the dynamic standard model are equivalent to equation (4) and (5) for the constant standard model. The model is different when $v_j = 1$, which is when the standard of accountability changes over time.

If θ_{it} was fully observed, then the likelihood functions above would be equivalent to independent ordinal logistic regression models. However, since this is not the case, all of the parameters of interest, the latent variable θ_{it}, the item difficulty cut-points α_{tjk} or α_{jk}, and the item discrimination parameters β_j, must be estimated simultaneously so that the model is identified. This issue necessitates the use of Bayesian simulation.

To estimate the models, I set the same priors on the latent variable estimate θ_{it} for both of the models compared in this paper such that $\theta_{it} \sim N(\theta_{it-1}, \sigma)$ for all i and t except when $t = 1$ and then

[18] For each item with dynamic difficulty cutpoints, $y_{itj} = k$ if $\alpha_{tjk-1} < \theta_{it}\beta_j + \varepsilon_{itj} < \alpha_{tjk}$, where ε_{itj} is an error term and $\alpha_{tj0} = -\infty$ and $\alpha_{tjk_j} = \infty$.

is $\theta_{i1} \sim N(0,1)$. This parameterization captures an old idea in the human rights literature: repression "radiates an after-life" which decreases the need for future repressive actions by the state for a certain period of time.[19] Both of the latent variable models formalize this idea and model comparison statistics help to validate it (Schnakenberg and Fariss 2012).

In both models, I specify $\sigma \sim U(0,1)$ to reflect prior knowledge that the between-country variation in human rights respect will be higher on average than the average within-country variance.[20] I specify the item-discrimination parameters $\beta_j \sim Gamma(4,3)$ to reflect the prior belief that all variables contribute significantly and in the same direction to the latent variable. These parameters estimate the strength of the relationship between values of the latent variable and the probability of being coded at a given level for one of the repression variables.[21]

For the dynamic standard model, I relax the assumption about the item difficulty cut-points made in other latent variable models and allow the α parameters to vary over time such that the priors of $\alpha_{tjk} \sim N(\alpha_{t-1,jk},4)$, subject to the ordering constraint that $\alpha_{tj1} < \alpha_{tj2} < \cdots < \alpha_{tjK-1}$ for all j. When $t = 1$ then $\alpha_{1jk} \sim N(0,4)$. By allowing the item difficulty cut-points for the standards-based variables to vary over time, I am able to assess how the probability of being coded at a specific level on the original ordered repression variables changes from year to year.[22] The priors for the α parameters for the event-based data in the dynamic standard model are $\alpha_{jk} \sim N(0,4)$, again subject to the same ordering constraint that $\alpha_{j1} < \alpha_{j2} < \cdots < \alpha_{jK-1}$ for all j. This is the same setup for all of the α parameters in the constant standard model. Table 4.2 summarizes the prior distributions for the model parameters and the differences between their implementation in the dynamic standard model and the constant standard model.

Table 1.3: Summary of Prior Distributions for Latent Variable and Model Level Parameter Estimates

Parameters	**Constant Standard**	**Dynamic Standard**
country-year latent variable (first year)	$\theta_{i1} \sim N(0,1)$	$\theta_{i1} \sim N(0,1)$
country-year latent variable (other years)	$\theta_{it} \sim N(\theta_{it-1},\sigma)$	$\theta_{it} \sim N(\theta_{it-1},\sigma)$
uncertainty of latent variable	$\sigma \sim U(0,1)$	$\sigma \sim U(0,1)$
event-based variable cut-points (constant)	$\alpha_{jk} \sim N(0,4)$	$\alpha_{jk} \sim N(0,4)$
standards-based variable cut-points (constant)	$\alpha_{jk} \sim N(0,4)$	—————
standards-based variable cut-points (first year)	—————	$\alpha_{1jk} \sim N(0,4)$
standards-based variable cut-points (other years)	—————	$\alpha_{tjk} \sim N(\alpha_{t-1,jk},4)$
slope	$\beta_j \sim Gamma(4,3)$	$\beta_j \sim Gamma(4,3)$

Note that both of these models are dynamic in the treatment of the latent variable θ

These models, like all item-response theory models, rest on an assumption of local independence. This assumption implies that any two item responses are independent conditional on the latent variable.

[19] See the quote by Duvall and Stohl (1983), which is cited by Stohl et al. (1986).

[20] This is not a consequential decision in terms of restricting the values of this parameter because the posterior estimates of σ from the converged model is less than 0.05, making the truncation decision unimportant.

[21] Prior sensitivity analyses suggested that this was not restrictive. When normal priors were specified for each β, the posterior densities rarely overlapped with zero. However, a model without this restriction is not identified with respect to rotation.

[22] Each model compared in this paper is 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).

This means that two item-responses are only related because of the fact that they are each an observable outcome of the same latent variable. There are three relevant local independence assumptions: (1) local independence of different indicators within the same country-year, (2) local independence of indicators across countries within years, and (3) local independence of indicators across years within countries. The third assumption is relaxed by incorporating temporal information into prior beliefs about the latent repression variable in both models and the changing standard of accountability in the dynamic standard model, which is captured by the item-difficulty cut-points.

Some readers may question the first assumption, which states that different repressive tactics are not causally related to one another within the same country-year but are instead only related through the underlying latent variable. This assumption is made implicitly in other projects that aggregate information about repression into one scale (Cingranelli and Richards 2012*a*,*b*; Gibney, Cornett and Wood 2012). See Jackman (2008) or van Schuur (2003) for further details about this assumption. More importantly however, different repressive tactics can be related to one another in theoretically important but non-causal ways. Fariss and Schnakenberg (2013), following many analysts before them[23], assume that repression is a useful tool for a leader because it produces the benefit of mitigating potential threats to the regime. However, the emphasis of this theory is that many repressive behaviors may be complementary policy options. A "complement" is defined if the presence of one repressive policy reduces the probability that another policy is made public, decreases the threat the first policy was used to address, or reduces the possibility of retribution faced by a leader caught using the original policy tool.[24] This theoretical distinction emphasizes the choices of the policy maker in selecting repressive tools. It is the leader selecting to both torture and imprison political opponents because of a threat, which is the underlying cause of repression generally and the two repressive behaviors specifically. This is an important theoretical and empirical issue that human rights scholars are currently grappling with (e.g., Fariss and Schnakenberg 2013; Conrad and Demeritt 2011). The model developed in this paper can be extended to help address this issue, however such an extension is outside the scope of this paper.

The models developed above assume that repression is caused by choices made by the regime that lead to some "true" level of repression in each country, each year. This is the latent variable of repression, which the models attempt to estimate based on observable outcomes. The observables are each a function of the latent variable, which are captured in the content of documents produced by human rights analysts working for Amnesty International and the US State Department. This content is then coded into data by research analysts (i.e., Cingranelli and Richards 1999, 2012*a*,*b*; Gibney, Cornett and Wood 2012; Hathaway 2002; Conrad, Haglund and Moore 2012). If the standard of accountability that the monitors use when assessing state behaviors changes over time, then data derived from these documents will be biased. Though the standard of accountability is not directly observable, I have parametrized it in the dynamic standard model. In the next section I compare the model estimates from this new

[23] See for example Carey (2006); Davenport (2007*b*); Mason and Krane (1989); Moore (1998, 2000); Poe and Tate (1994); Poe, Tate and Keith (1999); Poe (2004); Zanger (2000).

[24] Fariss and Schnakenberg (2013) find that, on average, physical integrity rights abuses are compliments with one another each year (1981-2006).

model to those from the constant standard model, which allows me to demonstrate that the standard of accountability has increased over time.

1.5 Results: Physical Integrity Practices Have Improved

A comparison of latent variable estimates from the dynamic standard model with those from the constant standard model provide strong evidence for a new view of repression: *physical integrity practices have improved over time*. Unobserved changes in the standard of accountability explain why average levels of repression have appeared to remain unchanged as the constant standard model would suggest. Differences in the average level of the latent variable estimates are displayed in Figure 1.9. For the constant standard model to be more consistent with reality and for this same pattern to obtain, the monitoring agencies would need to produce the human rights reports consistently from year to year *and* the producers of the event-based data would need to use a less and less stringent definition of repression in the assessment of these events over time. Neither of these alternative behaviors are supported by the theory nor the model comparison tests, which I introduce here.

I test to see if the dynamic standard model is a better approximation of reality relative to the alternative constant standard model. Readers should keep in mind that all existing models of repression — those based on information from the annual human rights reports — make the same assumption about a constant standard of accountability over time. Models of repression include all of the existing human rights scales that aggregate information about different rights abuses from the annual human rights reports (Cingranelli and Richards 1999, 2012*a*,*b*; Gibney, Cornett and Wood 2012; Hathaway 2002) in addition to the latent variable model recently developed by Schnakenberg and Fariss (2012) and the factor analytic method used by Landman and Larizza (2009). By comparing the estimates from the two latent variable models, I am able to test the hypothesis that an increase in the standard of accountability — the probability of observing and therefore coding a repressive outcome — increases over time for the repression variables derived from the human rights reports.

For these tests, I first present a statistic called the Deviance Information Criterion or DIC for short. This statistic provides information analogous to a penalized likelihood ratio test or simply the comparison of adjusted-R^2 statistics from competing models. For the DIC statistic, relatively smaller values indicate that a model explains more of the variance in outcome variables compared to an alternative model. Recall that the outcome variables in both models are the original repression variables. Next, I present results from posterior predictive checks. These tests simply compare model predictions of the original repression variables generated from the two competing latent variable models with the original repression variables themselves. The results suggest that the dynamic standard model again outperforms its competitor.

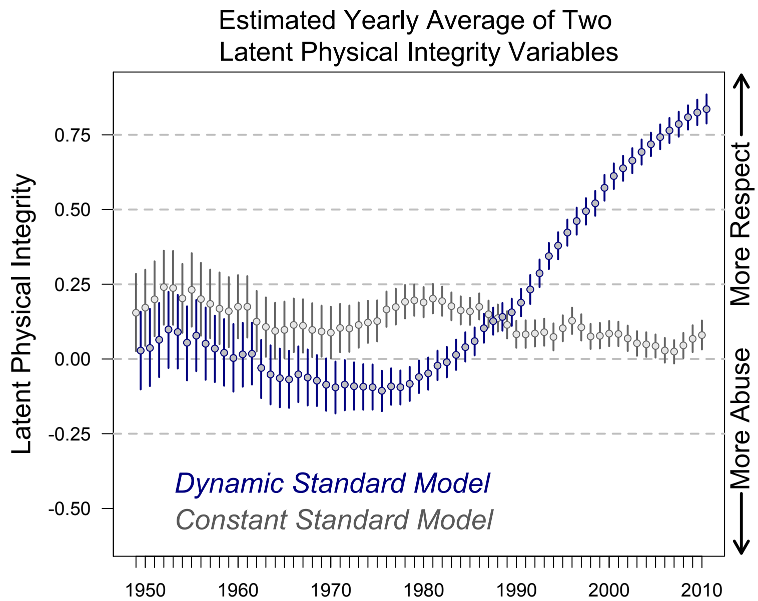


Figure 1.3: Yearly mean and credible intervals for latent physical integrity estimates from two models. The dynamic standard model allows the $k-1$ difficulty-cutpoints, the base-line probability of being coded at a certain level on the original standards-based repression variables, to vary over time. The standards-based variables are those which use human rights reports from the United States State Department or Amnesty International as their primary information source. The model with a constant standard estimates one set of $k-1$ cut-points for every repression variable including the standards-based variables. The difference in the two sets of estimates suggests that an increasing standard of accountability explains why the average level of repression has remained unchanged over time when the changing standard is not taken into account. By allowing this standard to vary with time, a new picture emerges of improving physical integrity practices over time, which begins after initially deteriorating from the beginning of the period until the late 1970s. Section 10.5 of the appendix contains selected country examples similar to this figure.

1.5.1 Model Comparisons: Deviance Information Criteria (DIC)

The DIC statistic is a method useful for comparing the models in this paper because it penalizes more complex models so that the more parsimonious one is favored, all else equal (Gelman et al. 2003). Thus a smaller DIC for this more complex model is strong evidence of its improvement over alternatives. Spiegelhalter et al. (2002) proposed that differences of greater than five or ten provide substantial evidence in favor of the model with the lower DIC. The DIC statistics are 53706 for the dynamic standard model and 55027 for the constant standard model, which is a difference of several thousand in favor of the dynamic standard model. See Section 10.8 of the appendix for more details.

1.5.2 Model Comparisons: Posterior Predictive Checks

Posterior predictive checks assess the quality of the model by direct comparison of model predictions from competing models. At every iteration of the MCMC algorithm the model parameters can be used to make a prediction of each of the observed repression variables. The better fitting model should on average generate predictions closer to the observed values of these variables when compared to similar predictions from a competing model (Gelman and Hill 2007).

Formally, for each draw from the posterior distribution I predict each of the j items y_{itj} for every country-year observation for which y_{itj} is observed. Since there are 1000s of draws from the posterior distribution, indexed by d, I am able to calculate the sum of squared differences of observed y_{itj} and the posterior predicted values $\hat{y}_{itjd}^{dynamic}$ from the dynamic standard model using the equation: $S_{itj}^{dynamic} = \sum_d (y_{itj} - \hat{y}_{itjd}^{dynamic})^2$ and likewise for the posterior predicted values $\hat{y}_{itjd}^{constant}$ from the constant standard model: $S_{itj}^{constant} = \sum_d (y_{itj} - \hat{y}_{itjd}^{constant})^2$

I have aggregated the sum of squared difference for each observation to compare values for the same observation from the competing models. These comparisons are captured in Figure 1.4 and a table in Section 10.9 in the appendix. Figure 1.4 displays the proportion of observations such that : $S_{itj}^{dynamic} <= S_{itj}^{constant}$, or in words, when the sum of squared difference from the dynamic standard model are less than or equal to the constant standard model for each country-year observation for all of the repression variable. Proportions closer to 1 indicate that the dynamic standard model outperforms the constant standard model at predicting the original repression variables. Proportions closer to 0 indicate the opposite. Proportions at 0.50 indicate that both models are predicting the items with about the same amount of error relative to each other. The proportions increase as the number of observations with a smaller sum of squared deviation increases when comparing the dynamic standard model and the constant standard model. The dynamic standard model does a much better job of predicting the original repression variables, especially the event-based variables. The improvement occurs for the event-based variables because of the temporal bias that exists in the standards-based variables. The constant standard model does not account for this bias, which reduces its ability to accurately predict the values of the event-based data not affected by the changing standard of accountability.

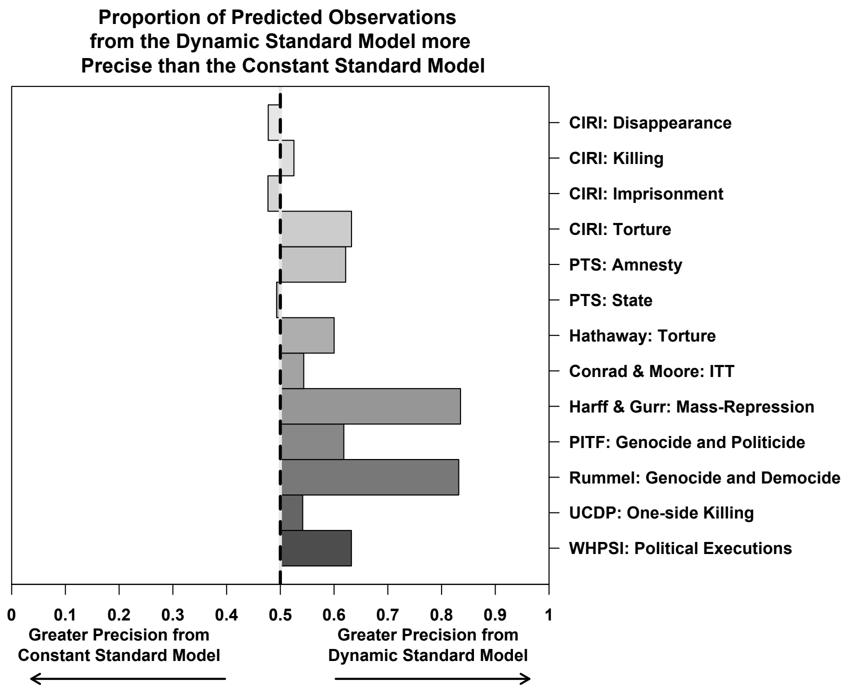


Figure 1.4: Proportions closer to 1 indicate that the dynamic standard model outperforms the constant standard model at predicting the original repression variables. Proportions closer to 0 indicate the opposite. Proportions at 0.50 indicate that both models are predicting the items with about the same amount of error relative to each other. The dynamic standard model does a much better job of predicting the original repression variables, especially the event-based variables.

1.6 Assessing the Changing Standard of Accountability

The model comparison statistics I presented above provide strong support for the new human rights data generated from the dynamic standard model but how does the changing standard of accountability influences the probability of being coded at a specific level of repression for the original standards-based variables over time? Figure 1.5 and Figure 1.6 presents panels that each display changes in the item difficulty cut-points (thresholds between values for each repression variable) from the dynamic standard model for each of the eight variables derived from the standards-based reports.

The changing standard of accountability does not affect all of the standards-based variables equally. Countries are far more likely to be coded for frequent torture based on the CIRI coding rules today than countries with similar levels of repression just a few decades ago. As the standard of accountability becomes more stringent, monitoring agencies look harder for torture, look in more places for torture, and classify more acts as torture. All of the standards-based variables with the exception of the CIRI imprisonment variable and the ITT torture variable are affected by changes to the standard of accountability (see Section 10.6 and 10.7 of the appendix for more details). However, as Clark (2001) discusses in her book, the original mission of Amnesty International was to document political imprisonment. The documentation of other human rights abuses came about as states responded to the advocacy efforts of Amnesty and other human rights NGOs. It is not surprising that the human rights reports consistently document political imprisonment over time. The lack of temporal change in the probability of coding levels of torture in the ITT data may reflect the relatively short period of coverage (1995-2005) or differences between Amnesty's Urgent Action Reports, which these data are based upon, and the annual report used by the other data sources. Additional analysis is necessary on this specific issue.

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.

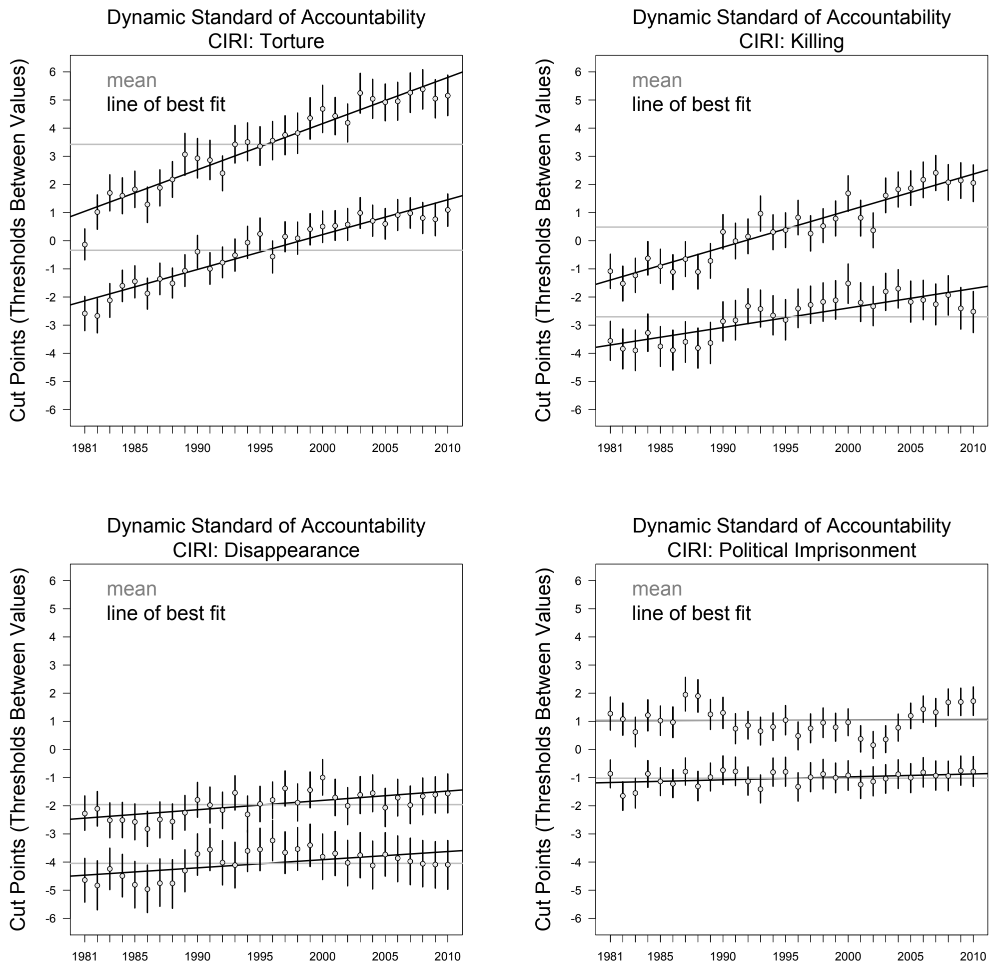


Figure 1.5: An increase in the difficulty cut-points translates directly into a statistically significant change in the probability of being classified as a 0, 1, or 2 on the original CIRI variables such that begin classified as 0 (e.g., frequent abuse) becomes more likely and 2 (e.g. no abuse) becomes less likely as a function of time. There is no statistical relationship between the Political Prison variable and time. See Section 10.7 of the appendix for the posterior estimates of these parameters and Section 10.6 for additional figures.

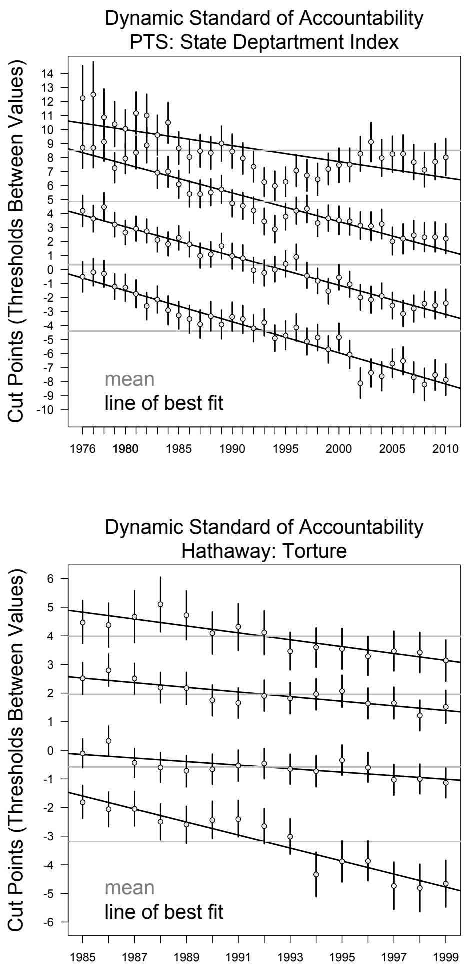


Figure 1.6: An decrease in the difficulty cut-points translates directly into a statistically significant change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original Political Terror Scale variables and the Hathaway Torture variable such that begin classified as 1 (e.g., little to no abuse) becomes less likely and 5 (e.g. widespread abuse) becomes more likely as a function of time. There is no statistical relationship between the ITT variable and time. See Section 10.7 of the appendix for the posterior estimates of these parameters and Section 10.6 for additional figures.

1.7 The Changing Standard of Accountability and Treaty Compliance: The Case of the Convention Against Torture

In this section, I illustrate the substantive importance of the changing standard of accountability for international relations theory by showing that ratification of the UN Convention Against Torture and respect for physical integrity rights is positive. This result contradicts negative findings from existing research. As the standard of accountability has increased over time, empirical associations with human rights data derived from standards-based documents and other variables will be biased if changes in the human rights documents are not accounted for. This is especially true for variables that measure the existence of institutions that are correlated with time such as whether or not the UN Convention Against Torture has been ratified.

In the international relations literature there are two opposing viewpoints on treaty effectiveness. Authors such as Morrow (2007), Simmons (2000), and Simmons and Hopkins (2005) argue that treaty ratification constrains actors to modify their behaviors by creating costs for noncompliance. An alternative viewpoint is that countries only ratify a treaty if they would have complied even in the absence of the treaty. Thus, treaties have no effect on the behavior codified within the treaty such as the level of cooperation (e.g., Downs, Rocke and Barsoom 1996; Von Stein 2005), or ratification of certain human rights treaties (e.g., Hathaway 2002; Hafner-Burton and Tsutsui 2005, 2007; Hafner-Burton and Ron 2009). The results presented here call into question this second viewpoint. The new latent variable model I have developed provides a way to improve the measurement of respect for human rights specifically and potentially the measurement of other forms of compliance in international relations more generally.

For this demonstration, I compare linear model coefficients using the dependent variable from the constant standard model and the dependent variable from the dynamic standard model. I estimate two linear regression equations using the latent physical integrity variables from the two measurement models. I regress these variables on a binary variable that measures whether or not a country has ratified the Convention Against Torture in a given year. I also include several control variables.[25]

New inferences are obtained by simply replacing the dependent variable derived from the constant standard model with the one from the dynamic standard model. Figure 1.7 plots the coefficient for CAT ratification from the linear models, which each use one of the two latent physical integrity dependent variables. The linear regression using the dependent variable from the constant standard model generates a negative coefficient, which corroborates results from earlier work. Comparison with the regression coefficient from the model using the dependent variable from the dynamic standard model is striking however. The coefficient has flipped signs and is statistically significant when compared with 0 ($p < 0.098$) and the alternative coefficient ($p < 0.004$). These results suggest that human rights protectors are more likely to ratify the treaty, that the treaty may in fact have some causal effect on human rights protection, or possibly

[25] I include measures of democracy (Marshall, Jaggers and Gurr 2013), the natural log of GDP per capita (Gleditsch 2002), the natural log of population, and the lagged value of the latent variable. See Section 10.12 of the appendix for more information about this and other specifications in addition to information about incorporating uncertainty inherent in the lagged latent variable.

both. Overall, these findings suggest that the treaty is not merely cover for human rights abusers.

Note that these models are not designed for causal inference and, though a variety of selection issues are known to exist when using this specification[26], the results from this type of model have spawned a large literature because of the counter intuitive, negative correlation found between ratification and respect for human rights (e.g., Hafner-Burton and Tsutsui 2005, 2007; Hathaway 2002; Hollyer and Rosendorff 2011; Vreeland 2008). Though this finding has been criticized (Clark and Sikkink Forthcoming; Goodman and Jinks 2003), it is generally taken for granted in the literature (Hafner-Burton and Ron 2009). Importantly, this new result calls into question a key assumption about state behavior made by several recent papers about human rights treaty compliance (e.g., Hollyer and Rosendorff 2011; Vreeland 2008).

Much additional testing is needed to probe the differences between existing empirical relationships and the new ones generated using the latent physical integrity estimates generated from the dynamic standard model. I am working on examining many other areas of human rights treaty compliance in another paper. The new model introduced in this paper might also be useful for analyzing other issue areas of treaty compliance in international relations, which I leave for future research.

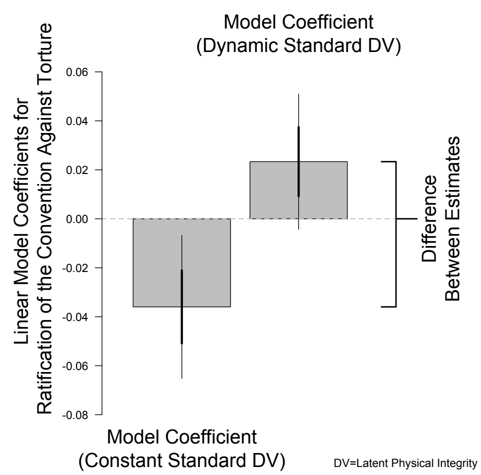


Figure 1.7: Estimated coefficient for CAT (the UN Convention Against Torture) ratification from the linear model using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error of the coefficient. The thin lines represent 2 $\pm$ the standard error of the coefficient. The difference between the coefficients is statistically significant ($p < 0.004$).

[26]See discussions in Neumayer (2005), Simmons and Hopkins (2005), Von Stein (2005), Simmons (2009), Hill Jr. (2010), and most recently Lupu (2013*b*). The selection issue that these author address is orthogonal to the differences in the two latent variable models. It is therefore sufficient to use this simple illustration to demonstrate how different inferences are obtained using the latent variable from the dynamic standard model.

1.8 Implications for Future Research

If changes in the standard of accountability are not addressed in applied research then biased inferences are the likely result. Figure 1.13 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 disagreement occurs because the dynamic standard model incorporates the changing standard of accountability, whereas the constant standard model, which is biased, does not.

The first option for analysts is to simply use the new latent repression estimates from the dynamic standard model. As I demonstrated in Section 1.7, a linear model can easily accommodate the latent repression estimates as the dependent variable. Schnakenberg and Fariss (2012) describe a method for incorporating the uncertainty associated with the latent variable estimates in this model or any other model that uses the lagged latent variable estimates as an independent variable (see Section 10.11 of the appendix for more details).

Analysts interested in any of the standards-based variables as a dependent variable should consider using a hierarchical model with the lagged estimate of repression generated from the dynamic standard model in addition to specifying time varying cut-points. This specification will help to avoid generating biased inferences. Through Bayesian simulations, programs such as JAGS, Stan, or WinBUGS can handle this more difficult to estimate model when using the standards-based variables. The alternative to this approach still involves specifying a time variable (a count of the number of years in the study beginning with the first year) interacted with the lagged repression estimates generated in this paper. I describe the specification for models using the original standards-based variables in the Section 1.8.1 below. I also present a procedure for modeling the original binary event data in Section 1.8.2. In both of these subsections, I present predictive validity statistics that corroborate the results from the DIC statistics and posterior predictive checks presented above.

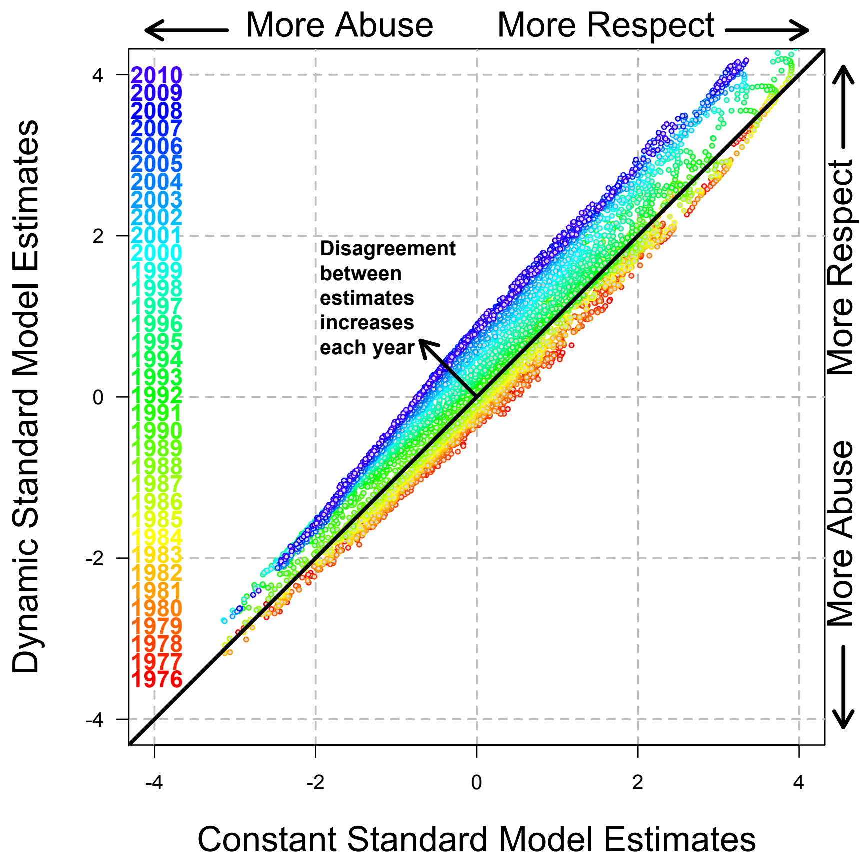


Figure 1.8: Relationship between the latent variable estimates generated from the Dynamic Standard Model on the y-axis and the estimates generated from the Constant Standard Model on the x-axis (1976-2010). The 45-degree line represents perfect agreement between the two estimates. Disagreement between the two sets of estimates increases as a function of time.

1.8.1 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 paper. 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 1.8.1). 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 simple 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.

Table 1.4: 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.

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]**

1.8.2 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 paper. I specify bivariate 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 1.8.2).

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 bivariate 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 1.8.2). 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.

Table 1.5: AUC statistics from bivariate 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 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.

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]

1.9 Conclusion

By allowing the standard of accountability to vary with time, a new picture emerges of improving physical integrity practices over time. Recall the research question I posed at the beginning of this paper: *Have levels of political repression changed?* To answer this question, I argued that the use of repressive policy tools appears unchanged over time because of the unaccounted for standard of accountability that monitoring agencies use to hold states responsible for abuse. I theorized that the standard of accountability, which I defined as a set of expectations or norms that state behaviors are measured against, changed over time because of the combination of several mechanisms. These mechanisms include improvements in the quality of human rights documents, increases in access to countries by NGOs that collect and disseminate information about repression, and finally changes in the subjective views of policy analysts on what constitutes a "good" human rights record. The theory allowed me to parameterize a measurement model of repression that incorporated the changing standard of accountability by allowing the baseline probability of observing a given level of repression for a specific repression variable to vary over time.

The results provide strong evidence that the changing standard of accountability affects the content of the human rights country reports produced annually by Amnesty International and the US State Department. The answer to the question posed above is that respect for physical integrity rights has improved after all. Put another way, the level of repression has decreased in the system over time, but this change was masked in the text of the human rights documents by a confounding factor not previously accounted for. By accounting for this additional factor, a new picture of global repression emerges in which conditions actually improve over the period of study (1949-2010) since hitting a low point in the mid 1970s. This result has implications for the research agenda of many human rights scholars and should not be left unaddressed in future research. In Section 1.7, I demonstrated that the empirical relationship between the new new physical integrity variable and ratification of the UN Convention Against Torture is positive, which contradicts results from earlier research. In Section 1.8.1 and Section 1.8.2, I demonstrated several methods for addressing the issue of temporal bias in empirical research that uses existing standards-based data and event-based data. The new latent variable model I have developed provides a way to improve the measurement of respect for human rights specifically and potentially the measurement of other forms of treaty compliance or other behaviors in international relations more generally.

To close, I wish to emphasize that the theory and model developed in this paper are not meant as a critique of any of the standards-based variables themselves. As should be clear, the theory and the model derived from it are focused solely on the changing standard of accountability, which influences the strategies used by monitoring agencies to generate primary source documents. It is these monitoring agencies and the documents they produce which are under investigation. In fact, this paper is itself a testament to the quality of the standards-based data because each of the variables included in the analysis reliably and accurately operationalizes content from these reports.[27]

The analysis conducted in this paper would not have been possible without all of the coding

[27] Each of the data sources report reliability statistics in their respective code books.

efforts made by the researchers associated with these data projects. The data accurately and reliably code both repression *and* the changing standard of accountability. However, the quality of inferences made about repression levels in different countries and over time depend on the clear specification of a theoretically informed model. By linking theory to model parameterization, it is now possible to measure this changing standard of accountability and, now that it has been identified, incorporate it into models of repression by using the estimates generated from this project.

1.10 Appendix

1.10.1 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 2012). 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.[28] 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.[29] Some scholars argued that event-based data 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.[30] 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

[28]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).

[29]The full wording of the PTS coding is below. It is taken from Gastil (1980). See also Gibney and Dalton (1996), Poe and Sirirangsi (1993), and Wood and Gibney (2010) for additional discussion of the development of these two indices.

[30]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).

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.

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 (2012).

- *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).

1.10.2 Event-Based Repression Variables

The event-based binary are variables drawn from Harff and Gurr (1988), Harff (2003), Rummel (1994*b*, 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 paper (Rummel 1966, 1994*b*, 1995; Rummel and Tanter 1974).[31] 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*b*, 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.[32] Several of the data sources that publish these binary variables also provide approximations of the number of events that occurred. 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.[33]

[31]The methods employed in this paper 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).

[32]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.

[33]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.

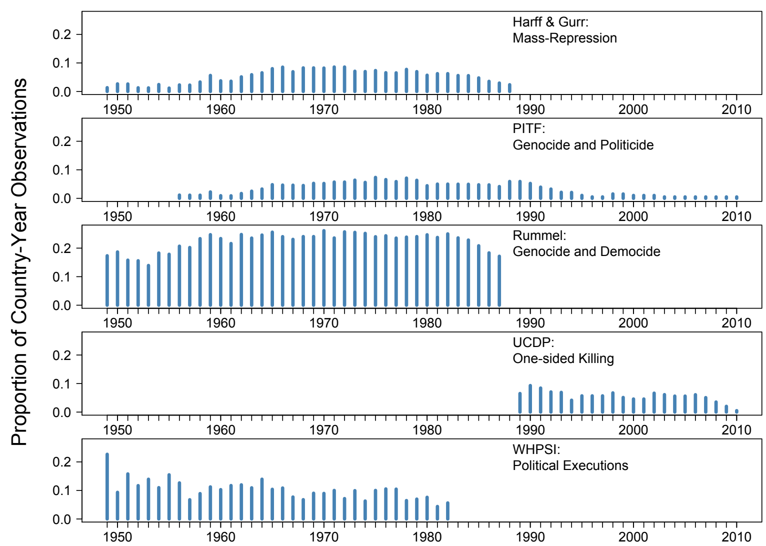


Figure 1.9: Event-based data over time.

1.10.3 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. 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 1.6. 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 paper.

Table 1.6: Changing information content in three human rights reports.

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

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 ... "

1.10.4 Additional Figures

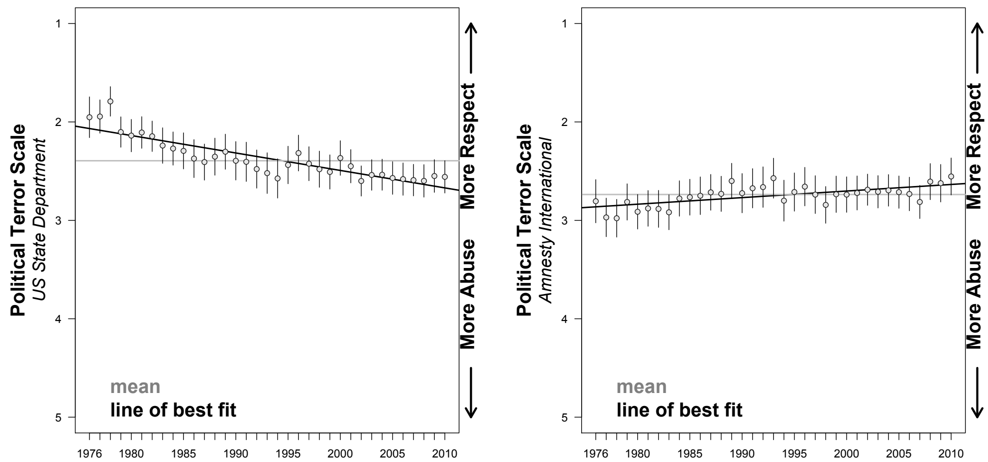
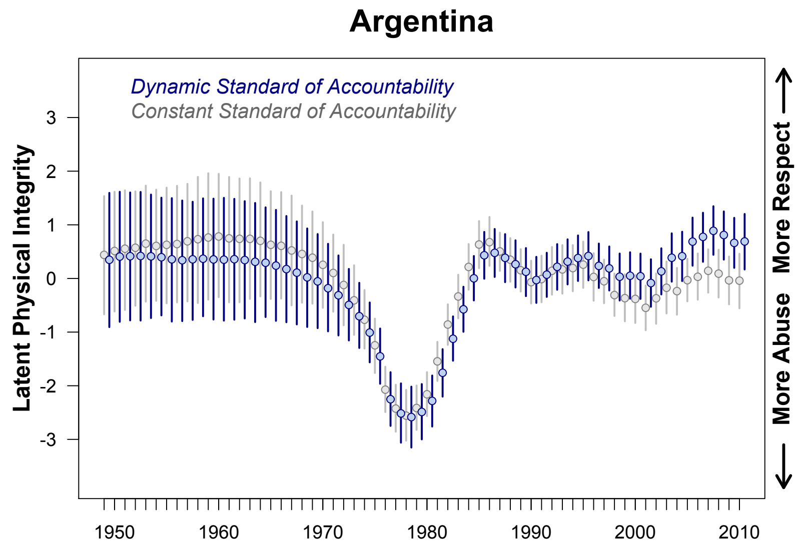


Figure 1.10: Yearly mean and 95% confidence intervals for the estimated level of repression using the Political Terror Scale index based on the US State Department reports (left), the Political Terror Scale index based on Amnesty International reports (right). Notice that the scales are inverted to be 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.

1.10.5 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.



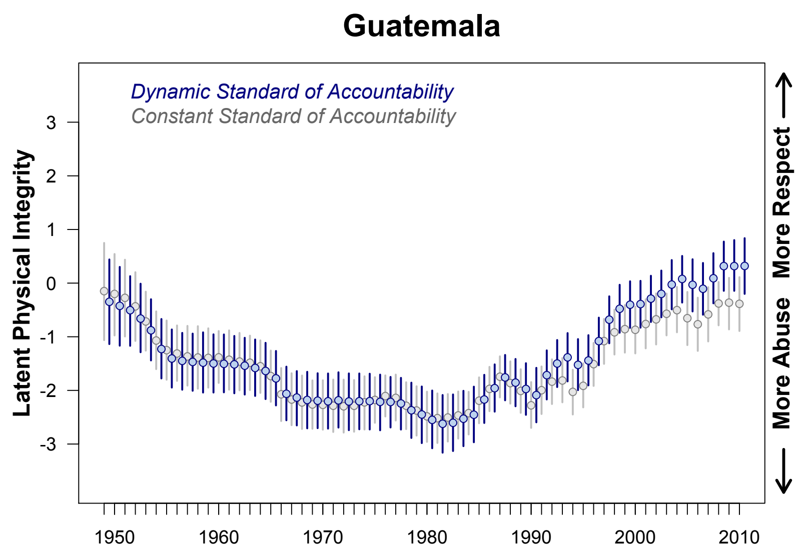


Figure 1.11: Selected country-year posterior estimates and credible intervals (1949-2010). Coverage extends back to 1949 because of the incorporation of multiple indicators of physical integrity rights violations. More information is available about state behavior in the post 1975 period so the estimates 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.

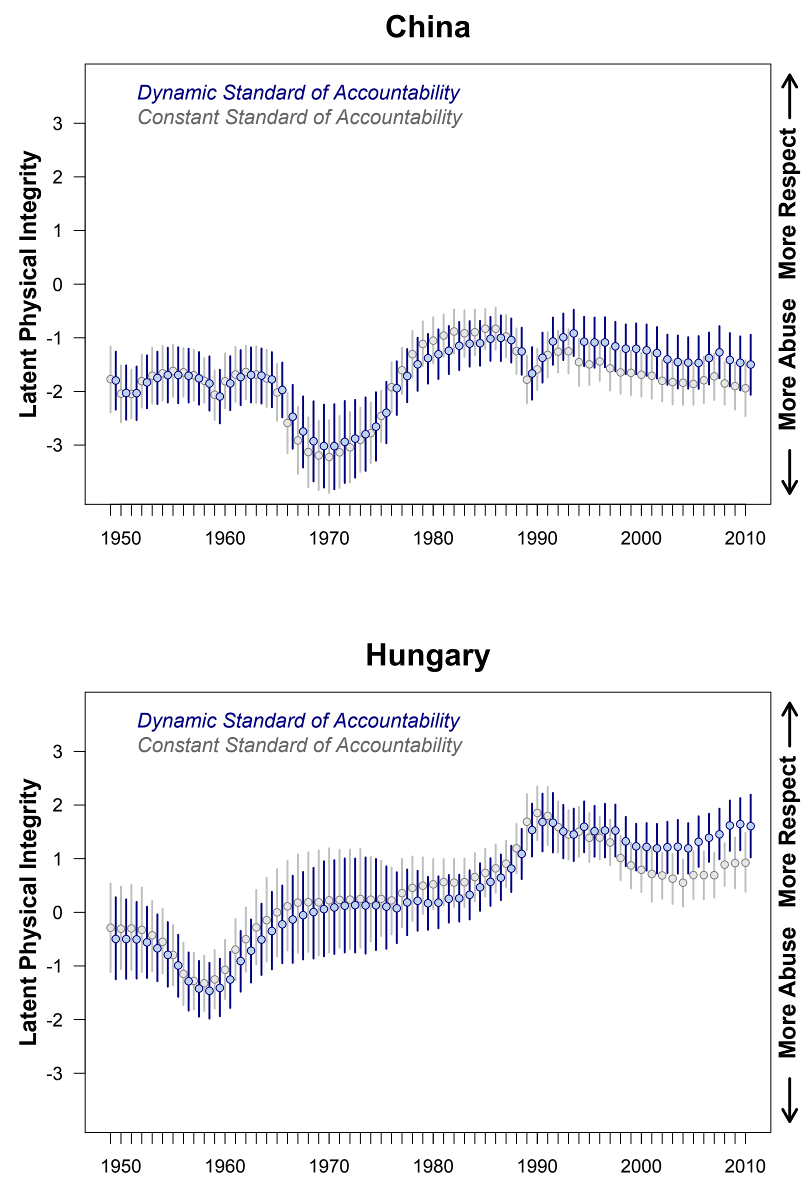


Figure 1.12: Selected country-year posterior estimates and credible intervals (1949-2010). The grey estimates represent those taken from the constant standard model. The blue estimates represent those taken from the dynamic standard model. The dynamic standard model explicitly accounts for changes in the standard of accountability over time. The difference between the two series increases as a function of time.

1.10.6 The Changing Standard of Accountability: Additional Plots and Tables

Figure 1.13 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 1.14, 1.15, 1.16, 1.17, 1.18, 1.19, 1.20, and 1.21. 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 1.10.6 and Table 1.10.6. Each row in the table reports slope coefficients and R^2 statistics from bivariate 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.

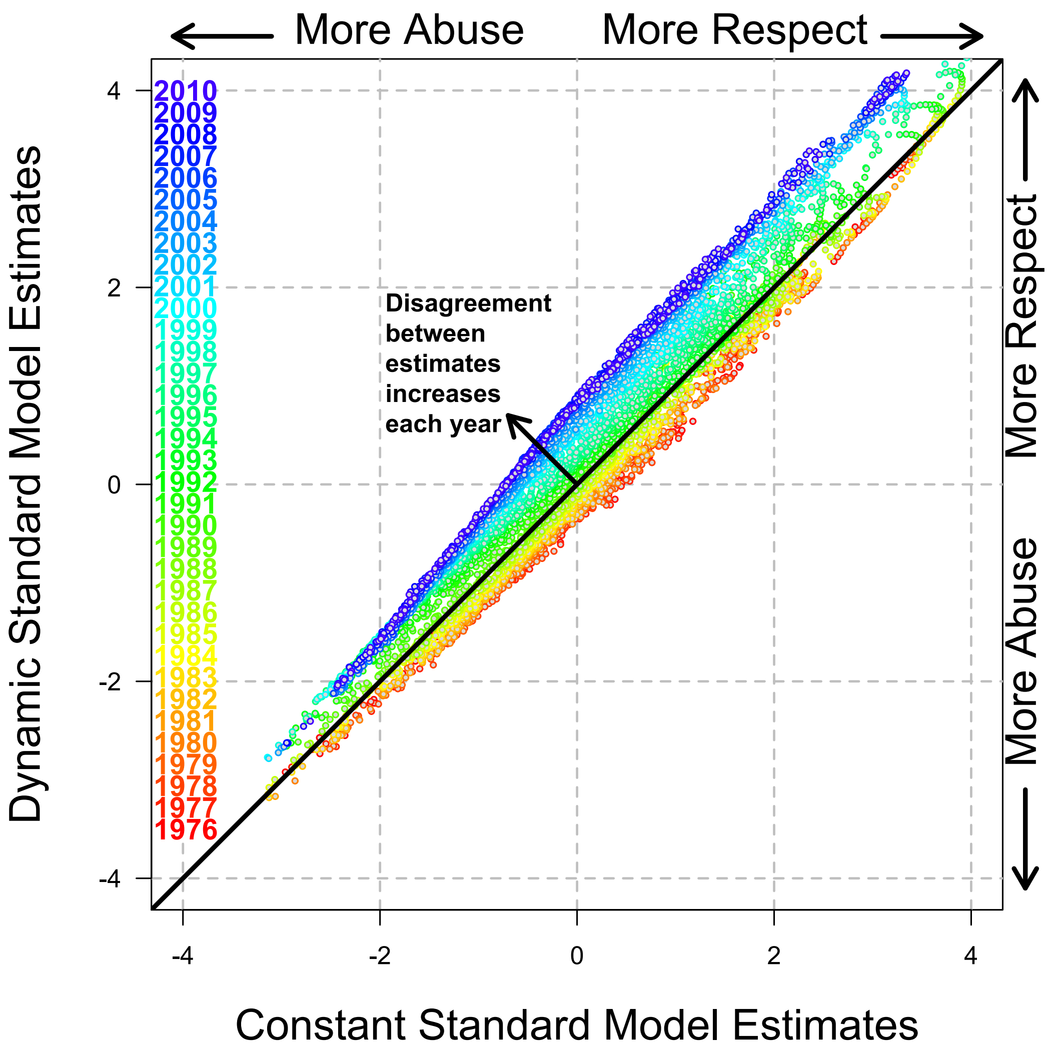


Figure 1.13: Relationship between the latent variable estimates generated from the Dynamic Standard Model on the y-axis and the estimates generated from the Constant Standard Model on the x-axis (1976-2010). The 45-degree line represents perfect agreement between the two estimates. Disagreement between the two sets of estimates increases as a function of time.

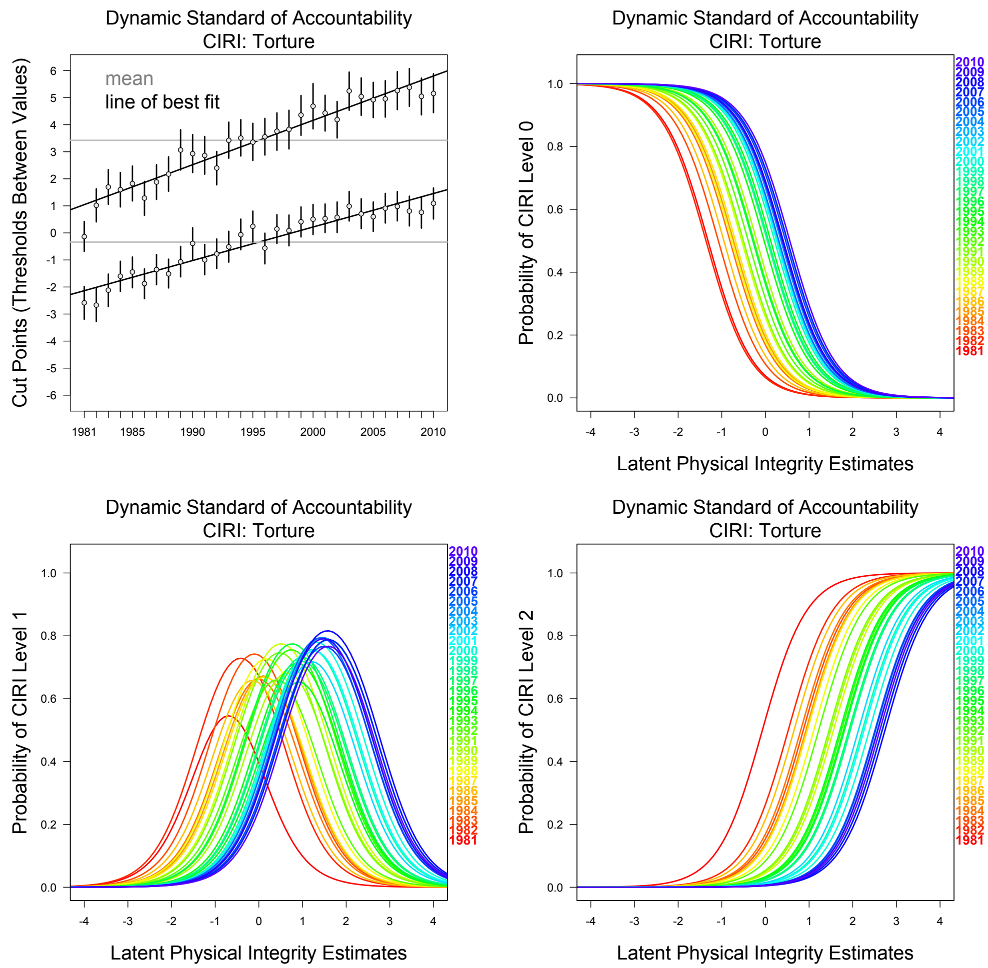


Figure 1.14: An increase in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 0, 1, or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent torture) becomes more likely and 2 (e.g. no torture) becomes less likely as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

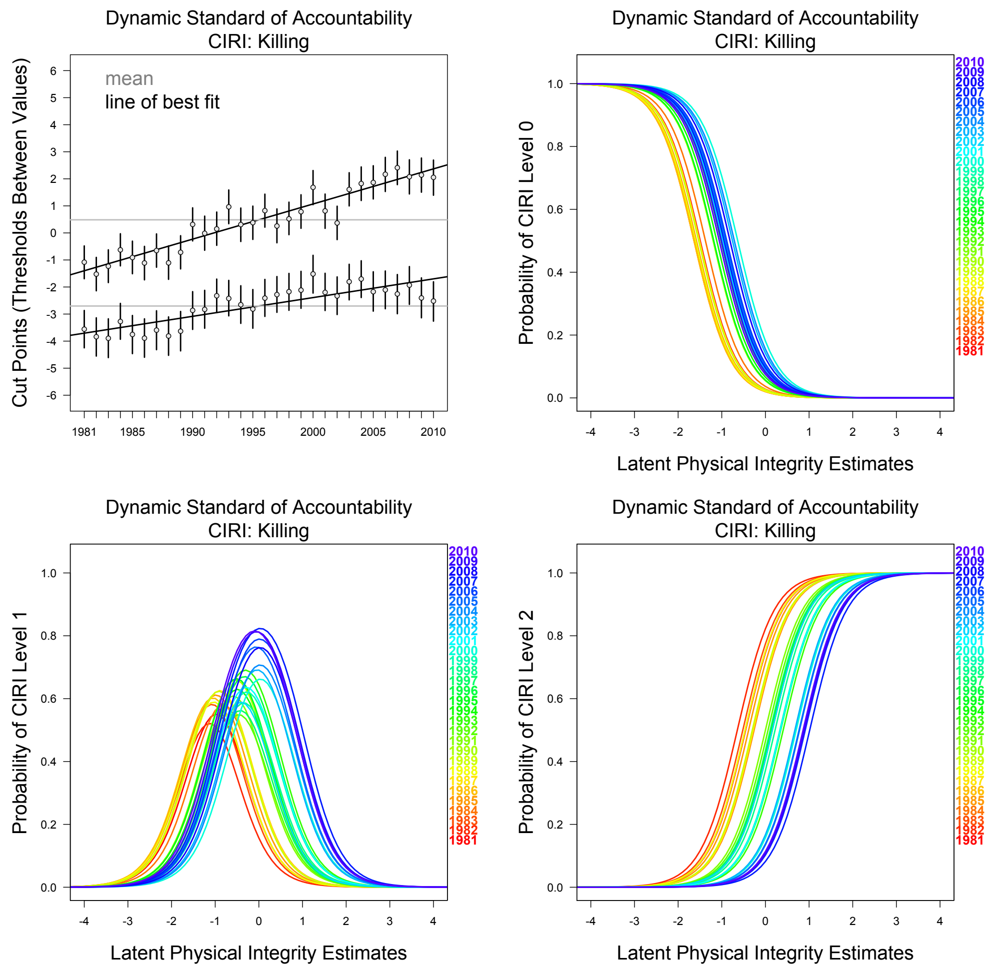


Figure 1.15: An increase in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 0, 1 or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent extrajudicial killing) becomes more likely and 2 (e.g. no extrajudicial killing) becomes less likely as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

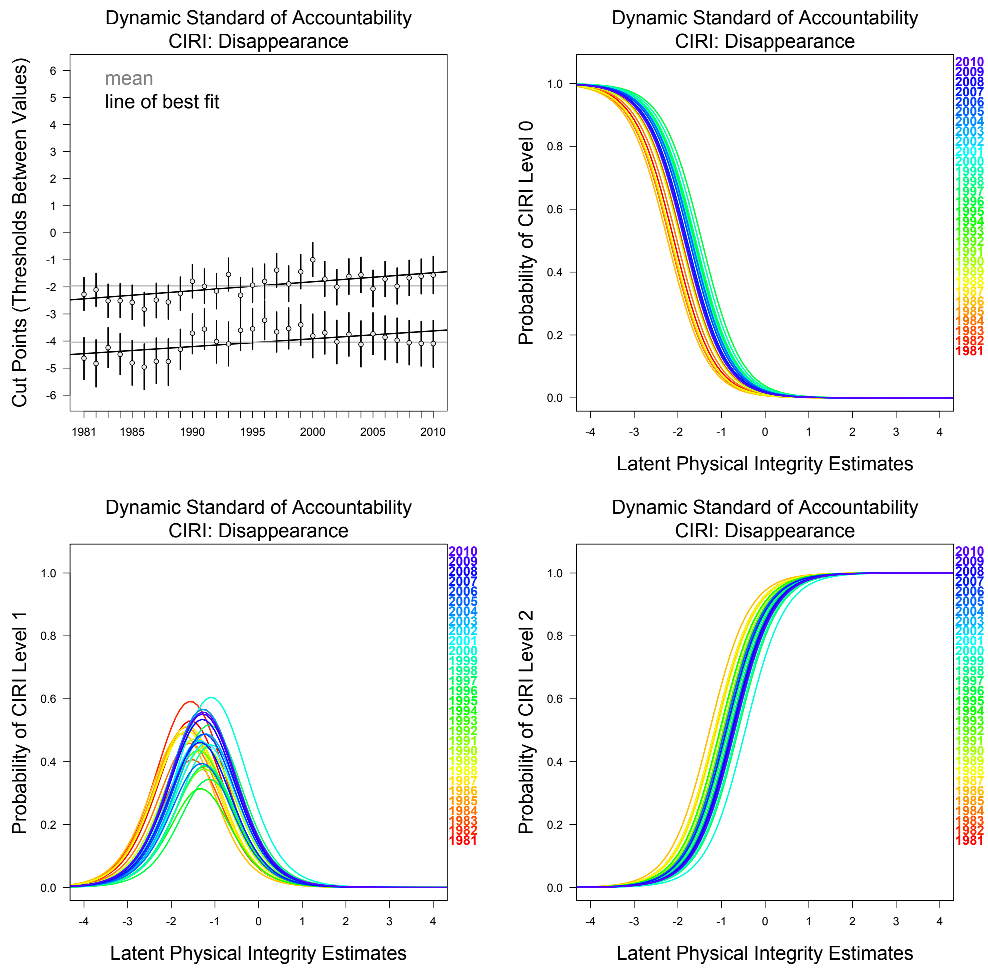


Figure 1.16: An increase in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 0, 1 or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent disappearances) becomes more likely and 2 (e.g. no disapperances) becomes less likely as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

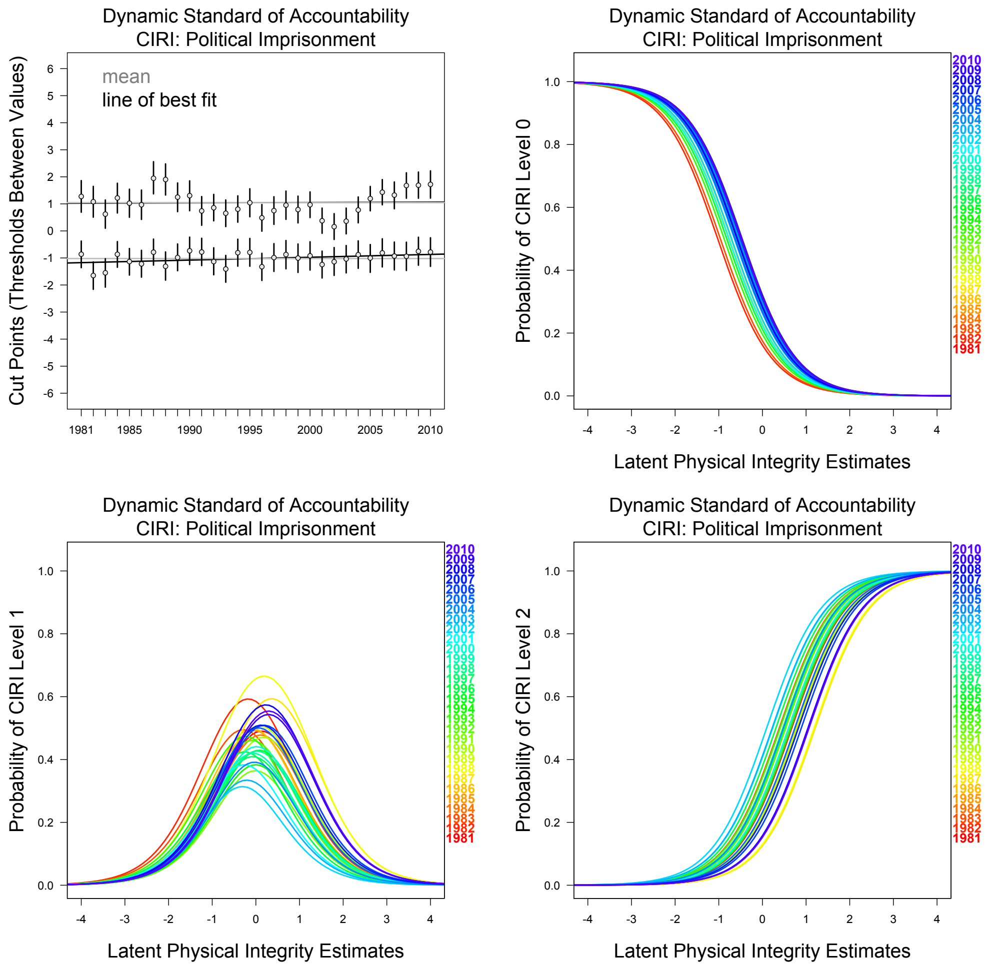


Figure 1.17: Very little change occurs over time for the difficulty cut-points in the upper left panel. Therefore, the probability of being classified as a 0, 1, or 2 on the original CIRI items such that begin classified as 0 (e.g., frequent political imprisonment) or a 2 (e.g. no political imprisonment) does not vary as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

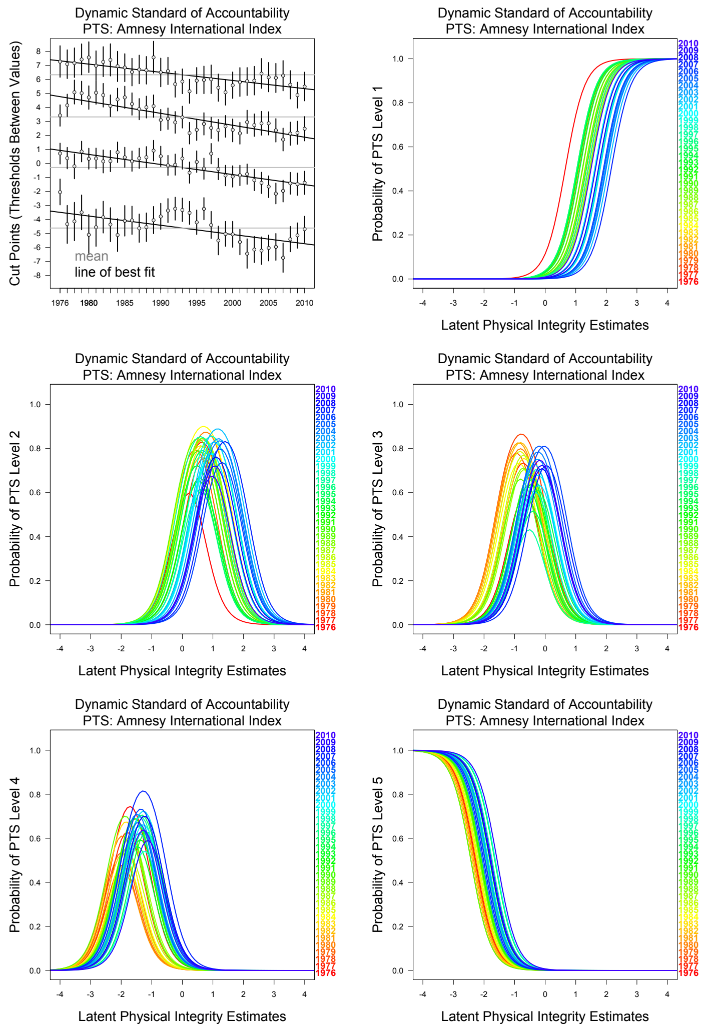


Figure 1.18: A decrease in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original PTS Amnesty Index such that begin classified as 5 (e.g., frequent abuse) becomes more likely and 1 (e.g. no abuse) becomes less likely as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

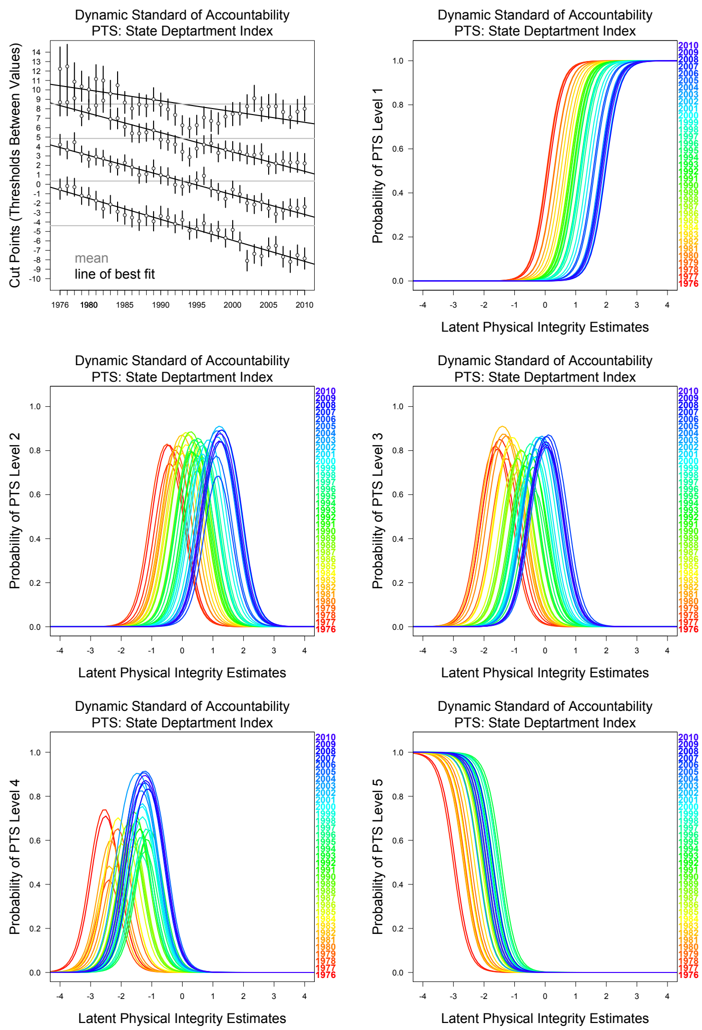


Figure 1.19: A decrease in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original PTS State Department Index such that begin classified as 5 (e.g., frequent abuse) becomes more likely and 1 (e.g. no abuse) becomes less likely as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

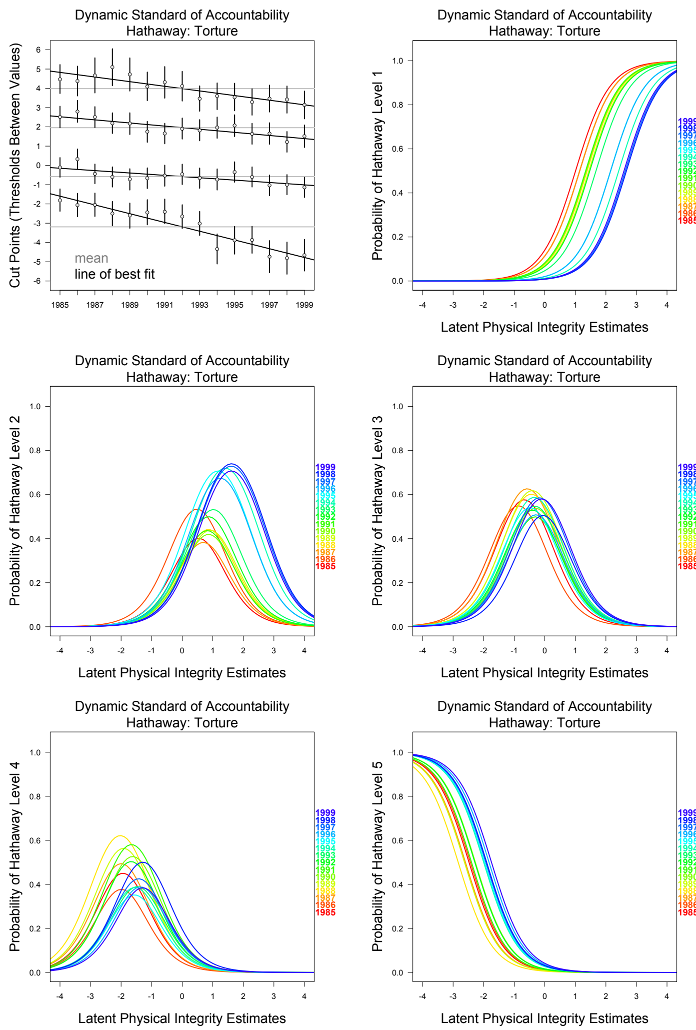


Figure 1.20: A decrease in the difficulty cut-points in the upper left panel translates directly into a change in the probability of being classified as a 1, 2, 3, 4, or 5 on the original Hathaway Torture Index such that begin classified as 5 (e.g., frequent tortyre) becomes more likely and 1 (e.g. no torture) becomes less likely as a function of time. See Section 1.10.7 for the posterior estimates of these parameters.

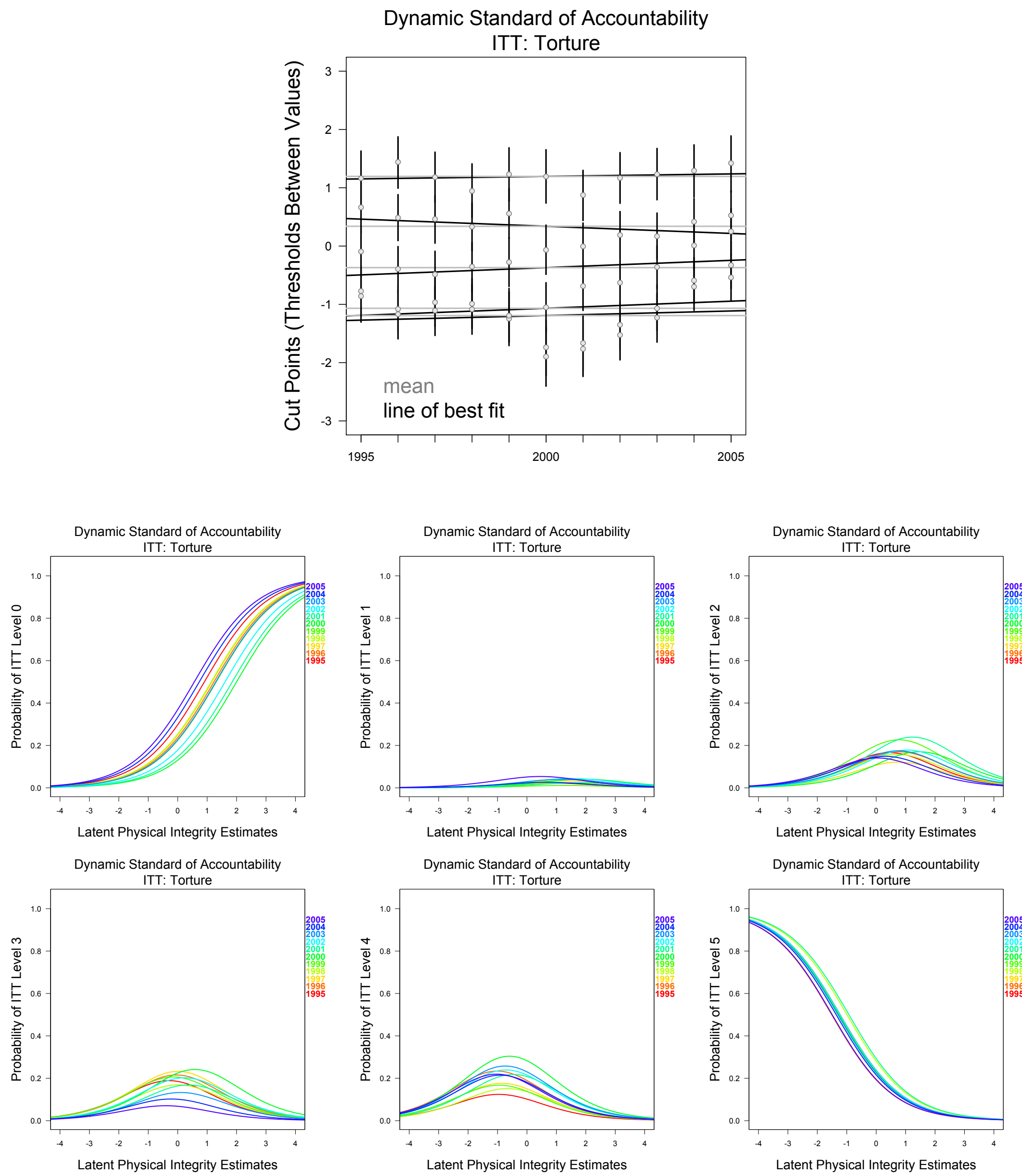


Figure 1.21: Very little change occurs over time for the difficulty cut-points in the upper. Therefore, the probability of being classified as a 0, 1, 2, 3, 4, or 5 on the original ITT Torture Index does not vary over time. See Section 1.10.7 for the posterior estimates of these parameters.

Table 1.7: Bivariate 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 bivariate 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 1.10.7 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]
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 1.8: Bivariate 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 bivariate 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 1.10.7 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]

1.10.7 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 1.10.6 and Table 1.10.6 above.

Table 1.9: Item difficulty posterior estimates and standard deviations for the CIRI disappearance item. These parameters are displayed visually in the main document.

t	$\alpha_{t,1}$	$\alpha_{t,2}$
1981	-4.641 (0.404)	-2.281 (0.309)
1982	-4.837 (0.437)	-2.116 (0.314)
1983	-4.237 (0.381)	-2.511 (0.316)
1984	-4.492 (0.393)	-2.508 (0.318)
1985	-4.819 (0.424)	-2.571 (0.325)
1986	-4.964 (0.423)	-2.825 (0.324)
1987	-4.754 (0.422)	-2.484 (0.327)
1988	-4.750 (0.435)	-2.564 (0.321)
1989	-4.304 (0.386)	-2.253 (0.316)
1990	-3.710 (0.378)	-1.790 (0.315)
1991	-3.558 (0.385)	-1.972 (0.325)
1992	-4.011 (0.404)	-2.142 (0.327)
1993	-4.109 (0.403)	-1.540 (0.309)
1994	-3.607 (0.373)	-2.306 (0.331)
1995	-3.555 (0.385)	-1.929 (0.320)
1996	-3.229 (0.359)	-1.794 (0.327)
1997	-3.661 (0.382)	-1.371 (0.317)
1998	-3.535 (0.384)	-1.892 (0.329)
1999	-3.398 (0.393)	-1.442 (0.317)
2000	-3.814 (0.410)	-1.011 (0.321)
2001	-3.684 (0.400)	-1.709 (0.328)
2002	-4.018 (0.417)	-1.993 (0.333)
2003	-3.749 (0.414)	-1.612 (0.325)
2004	-4.117 (0.435)	-1.545 (0.334)
2005	-3.732 (0.398)	-2.063 (0.352)
2006	-3.853 (0.423)	-1.720 (0.330)
2007	-3.975 (0.423)	-1.978 (0.337)
2008	-4.054 (0.423)	-1.668 (0.338)
2009	-4.090 (0.421)	-1.607 (0.329)
2010	-4.087 (0.443)	-1.568 (0.347)

Table 1.10: Item difficulty posterior estimates and standard deviations for the CIRI extrajudicial killing item. These parameters are displayed visually in the main document.

t	$\alpha_{t,1}$	$\alpha_{t,2}$
1981	-3.557 (0.360)	-1.080 (0.311)
1982	-3.844 (0.365)	-1.532 (0.311)
1983	-3.894 (0.365)	-1.233 (0.310)
1984	-3.272 (0.337)	-0.620 (0.303)
1985	-3.762 (0.361)	-0.916 (0.312)
1986	-3.896 (0.364)	-1.109 (0.313)
1987	-3.593 (0.363)	-0.655 (0.313)
1988	-3.812 (0.357)	-1.105 (0.313)
1989	-3.641 (0.372)	-0.717 (0.315)
1990	-2.855 (0.348)	0.315 (0.319)
1991	-2.824 (0.354)	0.000 (0.327)
1992	-2.326 (0.324)	0.143 (0.310)
1993	-2.430 (0.333)	0.968 (0.318)
1994	-2.654 (0.351)	0.310 (0.317)
1995	-2.818 (0.364)	0.378 (0.311)
1996	-2.412 (0.342)	0.834 (0.315)
1997	-2.284 (0.340)	0.261 (0.319)
1998	-2.165 (0.345)	0.526 (0.312)
1999	-2.120 (0.347)	0.788 (0.316)
2000	-1.511 (0.352)	1.677 (0.322)
2001	-2.197 (0.351)	0.812 (0.317)
2002	-2.332 (0.353)	0.372 (0.318)
2003	-1.798 (0.344)	1.607 (0.315)
2004	-1.696 (0.348)	1.832 (0.309)
2005	-2.169 (0.370)	1.859 (0.310)
2006	-2.110 (0.364)	2.172 (0.324)
2007	-2.256 (0.366)	2.415 (0.324)
2008	-1.932 (0.352)	2.079 (0.321)
2009	-2.404 (0.374)	2.141 (0.328)
2010	-2.517 (0.383)	2.054 (0.336)

Table 1.11: Item difficulty posterior estimates and standard deviations for the CIRI torture item. These parameters are displayed visually in the main document.

t	$\alpha_{t,1}$	$\alpha_{t,2}$
1981	-2.585 (0.308)	-0.136 (0.285)
1982	-2.675 (0.306)	1.029 (0.318)
1983	-2.118 (0.296)	1.704 (0.327)
1984	-1.598 (0.290)	1.591 (0.331)
1985	-1.439 (0.282)	1.822 (0.329)
1986	-1.870 (0.285)	1.289 (0.312)
1987	-1.353 (0.282)	1.885 (0.326)
1988	-1.510 (0.274)	2.176 (0.329)
1989	-1.066 (0.292)	3.062 (0.374)
1990	-0.390 (0.291)	2.941 (0.360)
1991	-0.993 (0.290)	2.863 (0.355)
1992	-0.775 (0.283)	2.397 (0.319)
1993	-0.520 (0.288)	3.424 (0.340)
1994	-0.060 (0.287)	3.513 (0.348)
1995	0.236 (0.290)	3.356 (0.347)
1996	-0.567 (0.289)	3.559 (0.348)
1997	0.137 (0.280)	3.757 (0.353)
1998	0.088 (0.286)	3.820 (0.360)
1999	0.413 (0.281)	4.359 (0.374)
2000	0.504 (0.277)	4.679 (0.425)
2001	0.525 (0.275)	4.436 (0.332)
2002	0.573 (0.284)	4.186 (0.351)
2003	0.994 (0.275)	5.247 (0.360)
2004	0.701 (0.279)	5.040 (0.354)
2005	0.603 (0.274)	4.928 (0.342)
2006	0.907 (0.277)	4.955 (0.338)
2007	0.990 (0.279)	5.258 (0.356)
2008	0.802 (0.283)	5.381 (0.359)
2009	0.761 (0.294)	5.055 (0.352)
2010	1.097 (0.295)	5.147 (0.371)

Table 1.12: Item difficulty posterior estimates and standard deviations for the CIRI political imprisonment item. These parameters are displayed visually in the main document.

t	$\alpha_{t,1}$	$\alpha_{t,2}$
1981	-0.867 (0.253)	1.280 (0.295)
1982	-1.648 (0.256)	1.080 (0.290)
1983	-1.550 (0.264)	0.620 (0.268)
1984	-0.864 (0.252)	1.218 (0.277)
1985	-1.133 (0.256)	1.022 (0.271)
1986	-1.220 (0.253)	0.969 (0.273)
1987	-0.788 (0.247)	1.944 (0.306)
1988	-1.306 (0.263)	1.904 (0.291)
1989	-0.985 (0.253)	1.253 (0.271)
1990	-0.738 (0.257)	1.306 (0.283)
1991	-0.780 (0.259)	0.747 (0.272)
1992	-1.138 (0.253)	0.855 (0.252)
1993	-1.412 (0.254)	0.649 (0.252)
1994	-0.811 (0.254)	0.805 (0.253)
1995	-0.796 (0.259)	1.042 (0.262)
1996	-1.317 (0.256)	0.495 (0.253)
1997	-0.986 (0.261)	0.753 (0.260)
1998	-0.865 (0.264)	0.960 (0.264)
1999	-1.003 (0.264)	0.788 (0.253)
2000	-0.922 (0.267)	0.972 (0.251)
2001	-1.239 (0.255)	0.377 (0.244)
2002	-1.145 (0.265)	0.154 (0.245)
2003	-1.032 (0.260)	0.360 (0.245)
2004	-0.879 (0.259)	0.776 (0.247)
2005	-1.008 (0.262)	1.193 (0.236)
2006	-0.815 (0.268)	1.431 (0.242)
2007	-0.924 (0.263)	1.324 (0.250)
2008	-0.937 (0.269)	1.676 (0.242)
2009	-0.751 (0.266)	1.684 (0.245)
2010	-0.782 (0.272)	1.716 (0.259)

Table 1.13: Item difficulty posterior estimates and standard deviations for the PTS Amnesty item. These parameters are displayed visually in the main document.

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

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

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

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)

1.10.8 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 paper 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 1.17.

Table 1.17: 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).

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

1.10.9 Posterior Predictive Checks: Additional Table

Table 1.18: 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.

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

1.10.10 Analyzing the Ordered Repression Variables: Additional Table

Table 1.19: Model deviance statistics from bivariate 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.

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]**

1.10.11 Method for Incorporating Uncertainty into a Model

Schnakenberg and Fariss (2012) 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).

1.10.12 Information About The Convention Against Torture Model Specifications

I use the method described in above in Section 1.10.11 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 1.20: 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
$ln(GDPper\ capita_{t-1})$	0.045	0.007

Table 1.21: 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
$ln(GDPper\ capita_{t-1})$	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

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

Table 1.22: 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
$ln(GDPper\ capita_{t-1})$	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 1.23: 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
$ln(GDPper\ capita_{t-1})$	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

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).

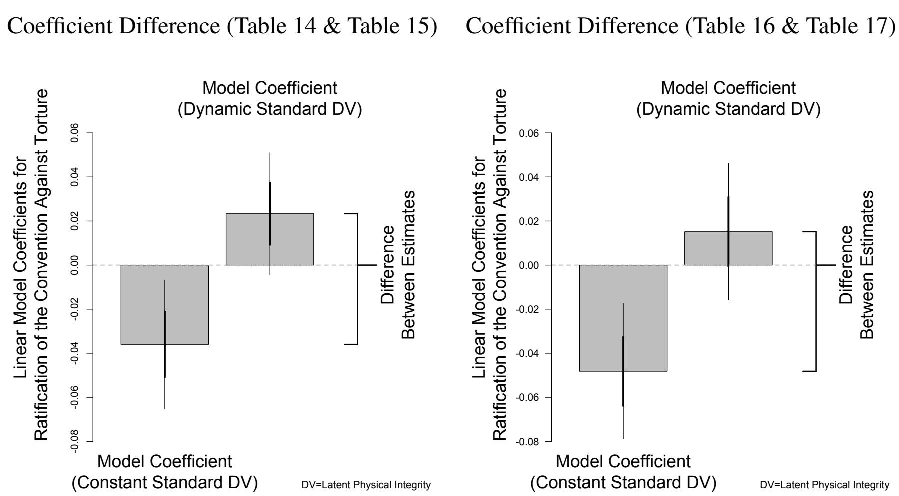


Figure 1.22: Estimated coefficient for CAT (the UN Convention Against Torture) ratification from the linear model using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error 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.

1.10.13 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)
 }
}
```

1.11 Acknowledgements

This paper was presented at the 2013 meeting of the Midwest Political Science Association and the 2013 meeting of the Western Political Science Association. I would like to thank the participants at these talks and also Megan Becker, Rob Bond, Chad Clay, Geoff Dancy, Jesse Driscoll, James Fowler, Blake McMahon, Micah Gell-Redman, Dimitar Gueorguiev, Jason Jones, Miles Kahler, David Lake, Yon Lupu, Jonathan Markowitz, Jamie Mayerfeld, Amanda Murdie, Maya Oren, Keith Schnakenberg, Paul Schuler, Brice Semmens, Jaime Settle, Neil Visalvanich, and Reed Wood for many helpful comments and suggestions. This chapter is currently under review at a peer-reviewed political science journal. The estimates from this paper along with the code necessary to implement the models in JAGS and R will be made publicly available at: `http://dvn.iq.harvard.edu/dvn/dv/HumanRightsScores`.

Chapter 2

Measuring Mutual Dependence Between State Repressive Actions

2.1 Introduction

> *Comparative quantitative assessment of human rights is hampered by the length of the list of internationally recognized rights. Not only is the list so long that it is hard to imagine gathering adequate data without an army of researchers (the International Human Rights Covenants contain more than thirty substantive articles, encompassing at least twice as many separate rights), but the results of such a comprehensive effort would almost certainly be overwhelming and bewildering in their complexity* (Donnelly and Howard, 1988: 214).

Over the last 20 years, scholars have compiled an impressive collection of human rights data (Carleton and Stohl 1985; Cingranelli and Richards 1999; Gibney and Stohl 1988; Gibney and Dalton 1996; Hathaway 2002; Poe and Tate 1994; Poe, Tate and Keith 1999; Richards, Gelleny and Sacko 2001). Though the need for data collection persists, sufficient progress has been made to allow researchers to address the rich complexity of this data. In this paper, we offer a simple tool to help understand the mutual dependencies between different human rights practices cross-nationally. This approach contrasts with most previous approaches, which assume either that rights are independent or that they are indicators of a single latent variable. We organize our inquiry around the following question: *how does the violation of many human rights influence the violation of a single right?*

Scholars in many fields are interested in the causes and consequences of human rights abuses; specifically the link between health and human rights (Leiter et al. 2006; Palmer et al. 2009; Singh, Govender and Mills 2007), the health effects of torture (Piwowarczyk, Moreno and Grodin 2000), the psychological causes (Fiske, Harris and Cuddy 2004; Smeulers 2004) and consequences (Silove 1999) of torture, and the political causes of human rights abuse (Cingranelli and Richards 1999; Keith 1999; Landman 2005; Landman and Larizza 2009; Poe and Tate 1994; Poe, Tate and Keith 1999; Powell and Staton 2009; Richards, Gelleny and Sacko 2001; Wood 2008). However, the research from these diverse fields do not directly assess the interdependent relationships among the rights that the Universal Declaration of Human Rights and the other international human rights treaties contain.[1]

Dependencies develop between different types of rights violations because repressive policy tools provide overlapping benefits to leaders and because repressive policies affect the costs of other repressive policies. The resulting decision-making by leaders should display common patterns of co-occurrence between different human rights violations. We contend that this pattern can be empirically modeled and then used to aid analyses of specific rights violations. We expect that a change in the costs of repression or the constraints on the use of repression should affect the pattern of rights abuses in a specific country and cross-nationally.

In this paper, we provide a general theory of interrelationships between state repressive actions and present a simple exploratory analysis designed to uncover mutual dependencies between human rights practices using graphical and statistical methods from network analysis (Wasserman and Faust 1994). Hu-

[1] For a complete discussion of the origins and definitions of all of the rights in the Universal Declaration of Human Rights see Donnelly (2003) and Donnelly and Howard (1988).

man rights scholars are aware of the important role that advocacy networks play in influencing country level rights practices.[2] Though we use similar tools, the goal of our paper is different. Instead of modeling NGOs or countries within a network framework we are modeling the rights themselves with these tools. The goal is not to characterize a literal network but to demonstrate how conceptualizing rights violations as nodes in a network leads to convenient graphical tools and data-reduction techniques that simplify an otherwise complex problem. The variables we derive allow for testing of hypotheses not typically considered by human rights scholars. We wish to emphasize that models that do not account for other human rights when a specific right is the dependent variable of interest will be theoretically under-specified. Our measurement strategy allows for researchers to focus on analyzing one level of one right while accounting for the mutual dependence of the other rights to that specific right of interest. Figure 2.1 diagrams this relationship.

In the remainder of this paper we define two idealized patterns of human rights abuse that emerge when governments make policy choices through (1) the simultaneous use of policy tools (complements) or (2) the replacement of one policy for another (substitution). To identify the conditions under which these theoretical patterns emerge and change we must first model the structure of the many interrelated human rights violations that occur across time and space. To accomplish this task we adapt a novel network model (Hidalgo et al. 2007) that links together several human rights variables (Cingranelli and Richards 1999; Richards, Gelleny and Sacko 2001) (nodes) based on the changes in the conditional probability (edges) of one right being violated given the violation of another right. The human rights network allows us to measure the position of a country as it moves towards violations of a specific right by providing a notion of distance from one bundle of practices to another. We then use the model to provide an initial assessment of likely sequences of human rights violations over time.

In this paper we focus primarily on describing the structure of the human rights co-occurance network and the variables derived from it. These new variables allow for the testing of many hypotheses related to the different types of relationships of human rights violations. To illustrate the potential of the network variables, we test for the human rights network influence on high levels (extreme) violations of four physical integrity rights and, using Monte Carlo simulations, we derive the step most likely to lead to the systematic use of actions that violate these rights. The result reveals that violations in the current year are strongly influenced by violations "nearer" to to that right in the human rights network and more weakly influenced by violations that are "farther" away. To conclude, we propose designs for additional tests of the relationships derived from the human rights network.

[2]See for example early theoretical work (Keck and Sikkink 1998; Korey 2001; Risse and Sikkink 1999) and more recent applied work that in some cases uses network analytic tools (Bell, Clay and Murdie 2012; Murdie and Davis 2012; Murdie and Bhasin 2011).

2.2 Conceptual Relationships Among Human Rights

Our theoretical approach assumes that repression is a result of cost-benefit analysis on the part of the leader. State leaders make policy decisions based on the costs and constraints associated with each policy choice. Some of these policy choices violate the rights of citizens. Repression is a useful tool for a leader because it produces the benefit of mitigating one of many possible threats to the stability of the regime (Carey 2006, 2007; Mason and Krane 1989; Poe, Tate and Keith 1999; Poe 2004; Zanger 2000). However, repression is potentially costly since the ruler can face retribution from local actors if the repression is made public.

Different repressive tactics can be related to one another in two ways. First, if two repressive tactics address the same type of threat to the regime, those tactics may be *substitutes*. In this case, an increase in the use of one repressive tactic reduces the need for the other. For instance, since extrajudicial killing and political imprisonment can both be used to eliminate influential anti-government activists, enhanced political imprisonment may reduce the number of killings and vice versa. However, since torture is a tactic designed for extracting information or intimidating individuals rather than eliminating them, one may not expect a similar substitution relationship betweeen torture and extrajudicial killing.

Second, if the presence of one repressive tactic reduces the probability that another tactic is made public or dampens the retribution faced by a leader caught using the tactic, those tactics may be *complements*. For instance, repressing journalists should reduce the probability that another repressive tactic is discovered, so we might expect increased censorship to be associated with increases in other rights violations. Furthermore, since all repressive tactics can extinguish retribution against the government, many repressive tactics should reduce the probability and magnitude of retribution for other repressive tactics.

The two theoretical relationships between different repressive tactics are not mutually exclusive. Thus, the relationship between two repressive tactics may be the product of countervailing forces. The relative importance of these two forces will determine the extent of the relationship between two tactics. However, we expect the complementary relationships between repressive tactics to be more common in practice than substitution relationships for two reasons. First, as described above, we expect some complementary relationship to be present among *all* pairs of repressive tactics since they all have the capacity to dampen retribution against the government. Second, substitution relationships may be more scarce since two repressive tactics are unlikely to serve exactly the same purpose. Though two tactics may have a similar benefit, the persistance of many different tactics suggests differences in the targets and situations calling for the use of each tactic. To the extent that complementary relationships between state repressive tactics are most important, different human rights violations should be expected to cluster in time and space. This hypothesis is more consistent with the high levels of correlation observed between many existing human rights indicators (Cingranelli and Richards 1999; Schnakenberg and Fariss 2012).

The clustering or complementary relationships between physical integrity abuses is well documented in the political science literature (Cingranelli and Richards 1999; McCormick and Mitchell 1997;

Poe and Tate 1994; Poe, Tate and Keith 1999); and the clustering of these policies is captured by the political terror scale (Gibney and Dalton 1996; Gibney, Cornett and Wood 2012; Wood and Gibney 2010) and the CIRI (Cingranelli and Richards 1999, 2012*a*) physical integrity index, which are used throughout the quantitative political science literature.[3] To be clear, these two scales only account for relationships between the four physical integrity rights; the right not to be tortured, imprisoned for political reasons, extrajudicially killed, or disappeared. The CIRI empowerment index (Richards, Gelleny and Sacko 2001) scales five additional rights; the right to free movement, free assembly and association, free speech, worker's rights and freedom of religion. However, to understand how the violation of one human right influence the violation of another right among many such rights we must think of each behavior as conceptually distinct and potentially heterogeneous in its relationship to each other right.

Our approach is theoretically linked to work on foreign policy substitution, which emphasizes the need to account for alternative policy options available to decision-makers when an existing policy becomes more costly.[4] However, though this literature emphasizes one particular relationship between policy options, in which policy-makers substitute one policy for another in response to new constraints, our theory emphasizes that many repressive tactics may be complementary policy options. When this is the case, we should expect the violation of one human right to increase when another right is violated. To summarize these relationships among each of the repressive tactics for which data is available, we conceptualize the system of relationships of human rights practices as a network of individual rights violations that can incorporate complementary and substitution relationships between repressive tactics.

2.3 The Human Rights Network

In social network research (Wasserman and Faust 1994), network models are constructed so that the nodes represent actors (e.g., friends, legislators) who are linked together by some relationship such as friendship or the cosponsorship of legislation (Bond et al. 2012; Christakis and Fowler 2008; Fowler 2006; Jones 2012; Jones et al. 2012; Settle, Bond and Levitt 2011). Recently, international relations scholars have begun to employ methods from the social network tool kit in order to examine the relationships that structure the international state system (Cranmer, Desmarais and Kirkland 2012; Cranmer, Heinrich and Desmarais Forthcoming; Kahler 2009; Lupu and Traag 2013; Maoz 2009; Murdie and Davis 2012). Scholars also use network methods to link together conceptual elements such as decisions from the Supreme Court of the United States, which are connected by judicial citations (Fowler et al. 2007; Fowler and Kam 2008; Lupu and Fowler 2012). This method has also been used to model the citation network of the European Court of Human Rights (Lupu and Voeten 2012).

[3]For reviews of the current state of the quantitative human rights literature see Landman (2004, 2005); for reviews of the early quantitative human rights literature see Poe (1990, 1991).

[4]For reviews of the foreign policy substitution literature see Bennett and Nordstrom (2000); Cioffi-Revilla and Starr (1995); Morgan and Palmer (2000); Moore (2000); Most and Starr (1984, 1989); Palmer and Bhandari (2000); Palmer, Wohlander and Morgan (2002); Regan (2000); Starr (2000). For reviews of the relationship between the literature on foreign policy substitution and the literature on human rights see Fariss (2010); Poe (2004); Rottman, Fariss and Poe (2009).

For our analysis of the relational structure of human rights violations we develop a conceptual network that links together human rights (nodes) with changes in the conditional probability (edges) of one human right being violated given the violation of another human right. We adapt our human rights network model from a model developed by Hidalgo et al. (2007) in which they analyze a network of export products linked together using a measure of conditional probability similar to the one we develop below.

Our model differs from the one in Hidaldo et. al. in some important ways. For instance, we choose a different definition of the connections between nodes and our application is considerably less complex. Both facts make our model simpler and easier to interpret. However, the novel insight that we borrow from Hidalgo et. al. is the use of network technology to analyze relationships between concepts rather than agents, countries or cases.

Characterizing the Human Rights Network

The human rights network is constructed using information about specific human rights practice as measured by the 13 CIRI human rights variables (Cingranelli and Richards 1999; Richards, Gelleny and Sacko 2001). The CIRI data include the four well-known physical integrity rights (the right to remain free from torture, political imprisonment, extrajudicial killing and disappearance)[5], the empowerment rights (the right to free association, a free press, free movement and freedom of religion)[6], the right to electoral self determination[7], and three variables that measure respect for women's political, economic, and social rights.[8] Each CIRI human rights variable measures the level of violation on an ordinal scale where, after reversing the scale, 0 indicates that the right is not violated, 1 indicates that the right is violated occasionally and 2 indicates that the right is violated frequently.

We have reversed the standard coding order from the original data in order to capture greater levels of human rights violations rather than greater levels of human rights respect.[9] From each of the 13 ordinal CIRI human rights variables we create two binary variables. The first measures if a moderate to extreme number of violations occurred, and the second measures only if an extreme number of violations occurred. Each of these variable pairs capture moderate to extreme human rights violations and extreme human rights violations respectively. We therefore create 26 new binary variables based on the 13 human right variables in the CIRI data set for 195 countries from 1981-2006. We use the network approach to derive a unidimensional measure of mutual dependence next and use that measure to construct empirically

[5]For a complete theoretical discussion of these rights see Carleton and Stohl (1985); Cingranelli and Richards (1999); Gibney and Stohl (1988); Gibney and Dalton (1996); Landman and Larizza (2009); Poe (2004); Poe and Tate (1994); Poe, Tate and Keith (1999); Poe et al. (2000).

[6]On empowerment rights see Richards, Gelleny and Sacko (2001).

[7]On the right to electoral self determination see Richards and Gelleny (2007*a*).

[8]On women's human rights see Poe, Wendel-Blunt and Ho (1997); Richards and Gelleny (2007*b*).

[9]Most of the CIRI variables are coded on a 3-point ordinal scale. Since it is necessary for our analysis that variables be on the same scale, we recode the three women's rights variables from a 4-point scale to a 3-point scale so that we can consistently compare each human right in the network. We do so for each of these variables by combining the two highest levels of respect (level 3 and level 2 into a single level 2 category). We make similar changes to the freedom of religion and freedom of movement variables which are dichotomous. For these variables we recode level 1 as level 2 and then reverse code the variable.

informed Monte Carlo simulations in the next sections of the paper.

With these 26 binary variables, we create a network variable measuring the probability of violating right i given the violation of another right j for all countries in a year t. Formally, we define the proximity as:

$$\phi_{i,j,t} = P(i=1|j=1) - P(i=1|j=0) \tag{2.1}$$

In words, the proximity between two rights is the change in the conditional probability of observing one right violated given the violation of another right. The proximity values are links that connect a group of hypothetical nodes used for illustrative purposes in Figure 2.2 and the human right nodes in Figure 2.3 (we describe both of these networks in detail below). The human rights network is a system-wide characteristic, therefore proximity values vary across years but not across countries in a given year.

We represent these new variables in an i-j-t array. That is, we generate a 26-by-26 adjacency matrix for each year t that we have data. Note, also that we set $\phi_{i,j,t} = 0$ when $i = j$.

Table 2.1: Adjacency Matrix of Proximity Values Between 26 Binary Human Rights Variables

$$\begin{pmatrix} \phi_{1,1,t} & \phi_{1,2,t} & \cdots & \phi_{1,26,t} \\ \phi_{2,1,t} & \phi_{2,2,t} & \cdots & \phi_{2,26,t} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{26,1,t} & \phi_{26,2,t} & \cdots & \phi_{26,26,t} \end{pmatrix}$$

We can characterize information in each of the adjacency matrices as displayed in Table 2.2. Positive values in each matrix indicate the complementarity between two right levels, such that the abuse of right level i is likely to occur contemporaneously with abuses of right level j. Negative values indicate that the two rights are substitutes, so abuse of right level i is negatively related to abuse of right level j. Table 2.2 summarizes the proportion of negative values that we observe for each year of human rights data. Note that on average, complementary relationships between violations of right levels occur with much greater frequency than substitutes in each year of the data. However, there are still several substitutive (negative) relationships that occur over time. On average 97.8% of the right-level-pairs are complements while 2.2% of right-level-pairs are substitutes. We wish to emphasize however, that these are system-year averages. Therefore there may be differences in the use of complimentary and substitutive policy combinations that vary based on country characteristics.

Substitutive relationships between the extreme levels of right-level-pairs are displayed in Table 2.7, which is located in the Appendix. The count is the number of years in which the particular right-level-pair is negative and therefore, representative of substitutive relationship. Notice that none of the pairs of substitutable rights are from the same CIRI Category as presented in Table 2.3. That is, none of the physical integrity rights are substitutes for any of the other physical integrity rights. Neither are any of the empowerment rights substitutes for any of the other empowerment rights. This pattern is consistent for the Women's right levels as well. Thus, the table is consistent with evidence that supports the use of the CIRI components to create the single dimensional physical integrity index and empowerment index

Table 2.2: Descriptive Statistics of Complementarity (+) and Substitution (-) Effects Between Repressive Actions

Year	Min ϕ	Max ϕ	Mean ϕ	Proportion of ϕ<0
1981	-0.064	0.795	0.290	0.018
1982	-0.074	0.809	0.281	0.028
1983	-0.095	0.830	0.273	0.034
1984	-0.140	0.826	0.274	0.017
1985	-0.106	0.805	0.275	0.022
1986	-0.207	0.881	0.281	0.034
1987	-0.137	0.803	0.274	0.040
1988	-0.210	0.825	0.274	0.049
1989	-0.077	0.813	0.295	0.031
1990	-0.063	0.778	0.295	0.003
1991	-0.016	0.823	0.315	0.003
1992	-0.223	0.828	0.296	0.022
1993	-0.209	0.848	0.272	0.015
1994	-0.236	0.868	0.276	0.049
1995	-0.101	0.845	0.286	0.015
1996	-0.053	0.858	0.295	0.012
1997	-0.015	0.834	0.320	0.006
1998	-0.014	0.860	0.327	0.003
1999	-0.196	0.873	0.310	0.012
2000	-0.054	0.863	0.316	0.012
2001	-0.234	0.876	0.306	0.025
2002	-0.497	0.882	0.303	0.028
2003	-0.154	0.869	0.311	0.022
2004	-0.178	0.856	0.288	0.025
2005	-0.221	0.890	0.284	0.031
2006	-0.186	0.874	0.253	0.028

that are often used in the literature (Cingranelli and Richards 1999; Landman and Larizza 2009; Richards, Gelleny and Sacko 2001; Schnakenberg and Fariss 2012).

Synthesizing information from the Human Rights Network

To reduce the human rights space to an easily interpretable unidimensional number we use the system-level proximity variable defined above to measure the total network influence on each right within the network. We define this concept as the *connectedness* of human rights around right i for each country k in each year t:

$$\omega_{i,k,t} = \frac{\sum_{j,t} x_{j,t}\phi_{i,j,t}}{\sum_{j,t} \phi_{i,j,t}} \tag{2.2}$$

Where $x_i = 1$ when a country violates right i and 0 otherwise. For example, the connectedness of a country to torture is the proportion of other rights that were violated in that year weighted by the

Table 2.3: CIRI Human Rights Variables

CIRI Category	**CIRI Variable**
Physical Integrity Rights	Disappearance Political Imprisonment Torture Extrajudicial Killing
Empowerment Rights	Freedom of Movement Freedom of Assembly and Association Freedom of Speech Worker's Rights Freedom of Religion
Electoral Rights	Electoral Self-Determination
Women's Rights	Women's Economic Rights Women's Political Rights Women's Social Rights

proximity of each right to torture in that year. Since the connectedness variable positions a country in the human rights network in relationship to a specific right i, values for $\omega_{i,k,t}$ are unique for each country k in each year t.[10]

A Hypothetical Network

Before describing the full network and connectedness variable, we illustrate the information that the connectedness variable captures with four hypothetical rights, A, B, C and D. Figure 2.2 represents one possible visualization of this network[11], which is generated by the hypothetical proximity values in Table 2.4.

Table 2.4: Simple Adjacency Matrix of Proximity Values Between Hypothetical Rights

$$\begin{pmatrix} \phi_{A,A} & \phi_{A,B} & \phi_{A,C} & \phi_{A,D} \\ \phi_{B,A} & \phi_{B,B} & \phi_{B,C} & \phi_{B,D} \\ \phi_{C,A} & \phi_{C,B} & \phi_{C,C} & \phi_{C,D} \\ \phi_{D,A} & \phi_{D,B} & \phi_{D,C} & \phi_{D,D} \end{pmatrix} = \begin{pmatrix} 0 & 0.6 & 0.9 & 0.3 \\ 0.6 & 0 & 0.2 & 0.1 \\ 0.25 & 0.2 & 0 & 0.3 \\ 0.3 & 0.1 & 0.4 & 0 \end{pmatrix}$$

Table 2.4 displays the proximity values that link the four hypothetical rights. As with the proximity values from the human rights network, these values are a system-wide characteristic and therefore vary across years but not across states in a given year. The proximity values thus capture the system wide change in the conditional probability of the violations of right i given the violations of right j. The connectedness value around a given right varies between 0 and 1. The connectedness of right violations

[10]Each value of $\phi_{i,j,t}$ is calculated for each year in the CIRI data set. Thus, $\omega_{i,k,t}$ is a unique country year value and differs for each individual right. For example the *connectedness* value will be different if the analyst's models extreme levels of torture as a dependent variable compared to another dependent variable such as extreme levels of political imprisonment. Finally, we note that the connectedness variable ω is not calculated with the ϕ value where $i = j$ so that the measure does not consider a right to influence itself.

[11]All graphs are generated using the Kamada and Kawai (1989) algorithm, which is implemented in the sna library (Butts 2012) in R (R Development Core Team 2011).

to right A in the simple network for some hypothetical state is determined by the number of other rights (B, C and D) that are violated.

Table 2.5: Proximity Values that Determine the Connectedness Around Right A

$$\begin{pmatrix} \phi_{A,A} & \phi_{A,B} & \phi_{A,C} & \phi_{A,D} \end{pmatrix} = \begin{pmatrix} 0 & 0.6 & 0.9 & 0.3 \end{pmatrix}$$

For example, the connectedness value around right A for state k that violates right C and right D (i.e., where $B_k = 0, C_k = 1$ and $D_k = 1$), is $\frac{(0*0.6)+(1*0.9)+(1*0.3)}{0.6+0.9+0.3} = \frac{2}{3}$. The most influential rights within the space are those with the highest proximity values as this illustrative case demonstrates. However, in order for the right to be of influence for a given state the right must be violated in that state. For example, the hypothetical state above does not violate right B. Thus, the proximity value that connects right B to right A is not used in the calculation of the connectedness variable. Finally, notice that the denominator in the connectedness equation above is the sum of all proximity values around right i, while the numerator is the sum of only those proximity values when country k is coded as violating right j in year t (when $x_{i,t} = 1$). Thus, the connectedness variable $\omega_{i,k,t}$ around human right i approaches 1 as the number of other human rights violations j increase in country k in year t.

Summarizing the relationships between rights

Figure 2.3 represents one of many potential visualizations of the human rights network. Since most of the relationships between the human rights variables are complementary, we use the graphs to investigate clustering of rights violations. First note that the visualization only contains 13 human rights nodes (extreme number of violations only) while the network is created using the 26 binary human rights variables defined above (moderate to extreme number and extreme number of violations). This simplification facilitates discussion of the network visualization but does not alter the operationalization of the connectedness variable or the inferences drawn from the Monte Carlo simulations discussed below. Each human rights variable acts as a node within the network. Each human right node is linked to every other human right node by a proximity value $\phi_{i,j,t}$. The plot is generated for all $\phi_{i,j,t} > 0.3$ for the average year, again to illustrate the emergent structure of the relationships inherent to the network (see Figure 2.4 for several network plots generated from alternative proximity thresholds). The connectedness variable however is operationalized to include all proximity values and thus information about the influence of the entire network on some right level i. The node sizes are proportional to $\sum_{j,t} \phi_{i,j,t}$ and represent the influence of one right on all other rights in the network. The arrows are directional information for $i \leftarrow j$ as $P(i = 1|j = 1) - P(i = 1|j = 0)$.

Other Approaches

The network approach described above is not a statistical model and is not meant to test the hypothesis that the rights are statistically related to one another. Rather, we have presented an exploratory tool for visualizing relationships between human rights practices. We then used this network information to develop a connectedness variable $\omega_{i,k,t}$ which can then be used to test hypotheses specifically about the interrelationship of rights abuses. Later in this paper, we provide an illustration in which this measure is used in a statistical model, but a few caveats are in order for such applications.

Some readers may notice an analogy between our network approach and other methods related to factor-analysis. Recently, scholars have used more sophisticated factor analytic methods (Cingranelli and Richards 1999; Landman and Larizza 2009; Richards, Gelleny and Sacko 2001) and item response theory methods (Schnakenberg and Fariss 2012) to better measure the clustering of human rights. Though these methods are similar in terms of the process of aggregating items into a coherent measure, the methods serve distinct purposes and have different implications for the types of hypotheses that can be tested.

A scholar using factor analysis or item response theory with the CIRI components would be modeling how each variable contributed to a latent level of human rights violations.[12] The network approach demonstrated in this paper serves a theoretically and methodologically distinct function when compared with this alternative approach. The latent variable approach assumes that the practices are indicators of a unidimensional latent variable and are independent conditional on the value of the latent variable. In contrast, the approach developed in this paper assumes that the practices are conceptually distinct but related to one another because of exogenous forces. It is worth noting that these alternative approaches are not easily testable against one another given the level of aggregation of currently available data. We consider this to be a promising and necessary avenue for future research and data-gathering efforts. In the remainder of this section we discuss the technique used in the original Cingranelli and Richards (1999) article in order to demonstrate how our approach is conceptually distinct.

Cingranelli and Richards (1999) investigate the scaling properties of the ordinal human rights variables using a technique called Mokken scaling (Mokken 1971). Mokken Scaling Analysis (MSA) can be described as a non-parametric item response theory model (van Schuur 2003) and is a stochastic version of a Guttman scale, in which items measure a single latent construct and can be ordered by difficulty (Guttman 1949).

Let θ denote a latent variable of interest. Though the researcher cannot observe θ, the researcher observes several items $1,2,...,J$. Let X_{ij} denote the score of subject i on item j, a random variable with realization $x_{ij} = 0,1,\cdots$. Also assume that each indicator has $m+1$ categories ($m = 1$ if the indicators are dichotomous, but this paper will focus on the case of $m > 1$). Since the values of the indicators are determined by the latent variable, the system can be characterized by the *item step response function* $P(X_{ij} \geq x|\theta)$ (Sijtsma and Molenaar 2002).

Mokken's model makes three important assumptions about the data. First, θ is a *unidimensional latent variable*, an assumption that can be tested using parameters from the MSA model (Cingranelli and

[12]The connection between these measurement models and the assumption of a latent variable giving rise to the indicators is more explicit in item response theory. The theory behind Principle Components Analysis, for example, is based on the atheoretical idea of simply finding a variance maximizing linear combination of the indicators. Thus, our comments in this section apply most directly to Mokken Scaling Analysis and other item response theoretic approaches. However, we note that authors who have applied methods such as Principle Components Analysis discuss them as if the first factor measures a unidimensional trait.

Richards 1999; van Schuur 2003). Second, the model assumes *latent monotonicity*, which means that the item step response function is strictly increasing on θ; $\theta_a \leq \theta_b \Rightarrow P(X_{ij} \geq x|\theta_a) \leq P(X_{ij} \geq x|\theta_b)$. Finally, the model assumes *local independence*, which means that the responses depend only on θ, $P(X_{i1} = x_{i1}, X_{i2} = x_{i2} \cdots X_{iJ} = x_{iJ}|\theta) = \prod_{j=1}^{J} P(X_{ij} = x_{ij}|\theta)$ (van Schuur 2003).

Mokken (1971) demonstrated that under the assumptions of a unidimensional latent variable, latent monotonicity, and local independence, the proportion of "correct" answers by subject i to item j is nondecreasing in the sum of all the items. These assumptions also imply that all of the items are nonnegatively correlated across all subsets of subjects (Mokken 1971). Under these assumptions the unweighted sum of the variables is nondecreasing in θ, a desirable feature of a measure.

Cingranelli and Richards (1999) utilize Mokken Scaling Analysis to confirm the scalability of the physical integrity rights indicators. This conclusion is valuable to the quantitative human rights literature because it validates the approach of using cumulative scales of disaggregated human rights variables. Furthermore, though previous approaches to quantitative human rights measurement assumed a unidimensional latent variable, the Mokken Scaling approach taken by Cingranelli and Richards (1999) allowed unidimensionality to be verified empirically.

Mokken Scaling Analysis and the other latent variable approaches (Landman and Larizza 2009; Schnakenberg and Fariss 2012) and the network approach demonstrated in this paper serve theoretically and methodologically distinct purposes. The network approach developed in this paper assumes that the different rights abuses are conceptually distinct but related to one another. This relationship is important if the researcher wishes to understand how some exogenous treatment affects both the right of primary interest and the other related rights.

The view that human rights behaviors arise from a single latent variable is, in most data, observationally equivalent to our current view that the concepts are conceptually distinct but complementary. However, we emphasize that the two models should be distinguished on the basis of usefulness for some particular purpose, rather than by truth value. The concepts of a "network" or a "latent variable" are simply useful abstractions for thinking about data and cannot be evaluated on the basis of truth. Our method is useful when interrelationships between human rights behaviors are of direct interest, and are not useful as an overall assessment of the latent level of respect for human rights in a country.

The example given early in this paper was that policy makers may cease violating a specific right after ratification of a UN human rights treaty but increase violations of some other rights. In this example, no change in the aggregate level of right violations may be observed. If this is the case then only the network approach developed in this paper will be able to test this hypothesis. We demonstrate the utility of the network approach with an analysis of extreme violations of four physical integrity rights in the next section of the paper. We have selected these variables to illustrate the potential of the network approach. Many additional hypotheses can be tested using this approach but are outside the scope of this paper.

2.4 Illustrations using physical integrity variables

In this section, we theorize about likely sequences of human rights violations with Monte Carlo simulations using our connectedness measure. This exercise is meant to illuminate the path a country might take from low violations of a particular right to high violations. Our approach is two-fold. First, for each physical integrity right variable, we use a logistic regression model to get a sense of the influence of the connectedness measure on occurrence of high-levels of violation of that right conditional on several covariates. Second, we use the logistic regression models and the co- occurrence networks to create Monte Carlo simulations which predict the step most likely to lead to extreme violations of four physical integrity rights.

The simulations rely on four logistic regression models, one for each of the four physical integrity rights. The dependent variable for each logistic regression is the presence of the most extreme level of violation of that right. The control variables used in the logistic regression models include *Gross Domestic Product (GDP) per capita*, *GDP per capita growth*, *Population Size*, *Population Growth*, *Level of Democracy*, *International war*, *Civil War*, *Military regime* and *British colonial legacy*. Since the main explanatory variable is lagged by one year, the control variables are also one year lags. These data are from Poe, Rost and Carey (2006) and detailed variable descriptions can be found in that articlie. We include a short description for each of these variables in the Appendix section of this paper. We have selected these variables to ensure that the simulations that we discuss next are generated using a plausible empirical model of human rights abuse. There are a number of additional variables that have been found to be related to human rights abuse.[13]

The main variable of interest in the regression models is the connectedness variable. Thus, the model assumes that the probability of observing an extreme violation of right i by country k in time t is

$$\begin{aligned} P(y_{i,k,t} = 1|\vartheta_i) &= \frac{1}{1+e^{-\vartheta_i}} \\ \vartheta_i &= \alpha_i + \beta_i \omega_{i,k,t-1} + \gamma_i M_{k,t-1} + \varepsilon_{i,k,t} \end{aligned} \tag{2.3}$$

where ω is the connectedness variable around the dependent variable, $y_{k,t}$. β_i is the parameter estimate of the relationship between connectedness and right level i. M is a vector of control variables lagged 1 year, which are described in the Appendix and γ_i is a vector of parameter estimates for these variables. In this exercise, we are interested in dynamics rather than simply co-occurrences, so we use a one-year lag of the connectedness variable to see if states that are "closer" in the network to a right violation in one year are more likely to violate the right in the next year. Finally, to further address dynamics, we use a cubic spline (Beck, Katz and Tucker 1998) or a cubic polynomial (Carter and Signorino 2010) to control for temporal

[13] See for example Davenport (2009); Davenport and Armstrong (2004).

dependence in the model. We estimate the model with both types of temporal variables but only display the results with the splines below since the substantive conclusions are very similar using either method. We run our statistical models in R (R Development Core Team 2011) using the Zelig library (Imai, King and Lau 2007) for all country-years between 1981 and 2006.[14]

The full parameter estimates from the logistic regression models are displayed in Table 2.6. The connectedness variable strongly predicts future extreme violations in all four logistic regression models. To illustrate this effect, Figure 2.5 displays 99% confidence intervals for the probability of extreme violations of each right at various levels of network connectedness. Moving from one standard deviation less than the mean connectedness score around torture, for instance, to one standard deviation greater than the mean results in a 112% increase in the probability of extreme violations of torture. Note that these effects incorporate heterogeneity of influence despite using a unidimensional measure, so we predict a higher likelihood of high violations of a right when countries are violating "nearer" rights, even holding constant the number of other rights being violated.

To derive the step most likely to lead to the extreme violation of one of the four physical integrity rights, we conducted Monte Carlo simulations on a counterfactual data set in which non-human-rights variables were held constant at their means and human rights variables were randomly sampled from the set of all permutations of human rights scores. This methods allows presentation of probabilities of the four physical integrity variables based on the distance of the nearest right that was violated in the previous year. The method of simulation are commonly used in political science, and described in King, Tomz and Wittenberg (2000).

Figure 2.6 shows the simulated probabilities of extreme violations of each right as a function of the "nearest" violated right in that country in the previous year. The simulations reveal differences in the probability of extreme level of each physical integrity right when a nearby right is violated as opposed to a right that is farther from it in the network. For example, when a country engaged in extrajudicial killing (at the "moderate to extreme level") in the previous year, the probability of extreme violations of torture was 0.28, in contrast to a probability of 0.18 when the nearest violated right is freedom of association. In contrast, the step most likely to lead to the extreme violation to political imprisonment was violations of rights to freedom of movement, a variable usually not considered to be derived from the same latent trait as the physical integrity variables. The pattern observed for political imprisonment contrasts with the view, found in Cingranelli and Richards, that sequencing of human rights violations proceeds in a simple fashion through physical integrity rights as a result of latent human rights levels. The sequence leading to imprisonment appears to rely on a simple conceptual relationship between the rights; political imprisonment is the mode of enforcement for violations of rights to movement or freedom of association. This relationship corroborates the network visualizations displayed in Figure 2.3 and Figure 2.4 in which the political imprisonment node connects the physical integrity rights abuses with the

[14]Each of the variables in the statistical model contained missing values. Missing values were imputed using Amelia II (King et al. 2001). We also include several additional variables to improve the imputation model. We include the POLITY IV data version 2006 (Marshall, Jaggers and Gurr 2003) and the Correlates of War Composite Index of National Capability (CINC) data version 3.02 (Singer 1987; Singer, Bremer and Stuckey 1972).

empowerment abuses.

2.5 Conclusion

In this paper, we have presented a theory of interdependence between human rights behaviors and illustrated that theory with data using a network approach that allows for the measurement, visualization and statistical analysis of the mutual dependencies between different repressive tactics. Our analysis suggests that rights violations are generally likely to co-occur and that the system of co-occurrence can be usefully represented in a low-dimensional measure. For instance, the measure can be used to illustrate how the bundle of human rights violations in a country influence likelihoods of different physical integrity abuses. For example, states that broadly violate “nearer” human rights are more likely to start torturing and less likely to quit. The simulation analysis empirically demonstrates the step most likely to lead to the wide spread use of four physical integrity rights.

The goals of this paper are primarily exploratory, and we hope the paper inspires more systematic and detailed exploration of relationships between various rights violations. For instance, although we have provided a general framework to explain relationships between rights, we have not applied the framework to give more specific predictions for specific pairs of rights violations. We consider this to be an important and exciting area for future research. Furthermore, for reasons of simplicity and illustrative value, we have not attempted a sophisticated statistical treatment of the problem of relationships between rights and have not presented many formal hypothesis tests. More sophisticated multivariate statistical models and structural equations models building off of the approach developed in this paper could be used to analyze these relationships. Finally, our network measure was constructed to be a system-wide measure in each year, although it is possible that patterns of co-occurrence of human rights violations vary considerably based on country characteristics. Though we consider the system-level variable to be intrinsically interesting as an analytical tool for characterizing repressive tools, it is straightforward to repeat our analysis on different subsets of countries for comparison.

The measures developed in this paper will be of both theoretical and methodological use to scholars conducting empirical analyses of human rights practices. Scholars frequently analyze the correlates of a particular human rights practice by considering some treatment of interest and a set of control variables. Just as frequently however, these scholars do not include other human rights practices on the right-hand side of the equation. These relationships are not only theoretically interesting, but may be important omitted variables in studies that focus on the violation of one particular right.

Furthermore, the insights from our analysis will also likely be of use to scholars interested in the effects of human rights abuse on human health and well-being. Our results suggest that isolating the effects of torture may be a difficult endeavor since individual subjects who experience extreme levels of torture are likely to have also experienced other types of human rights abuse (Silove 1999). Scholars should therefore account for other human rights that likely precede violations of torture such as political imprisonment, extra-judicial killings, and limitations on freedom of movement in the locations that they study.

Also, the human rights network may condition the effect of interventions (such as “naming and shaming”) meant to improve human rights practices (Demeritt 2012; Keck and Sikkink 1998; Krain 2012; Risse and Sikkink 1999; Meernik et al. 2012; Murdie and Bhasin 2011). Interventions aimed at preventing torture may be more effective when fewer violations of other rights are present, and ineffective when human rights are broadly violated. Researchers interested in these interventions may test interactions between our measure of network connectedness and their treatments of interest. Similarly, agencies may choose to devote resources to interventions with higher probabilities of success by focusing on countries with a few bad practices where human rights are otherwise generally respected. Analyzing such interventions by matching on previous values of network connectedness is one efficient way to control for these selection effects.

As a whole, the quantitative human rights literature will benefit from further examination of how human rights practices are related to one another by causal factors. It is our hope that other scholars will begin to account for the relationships that exist between the many different human rights.

	Prison	Torture	Killing	Disappear
Intercept	-4.37 *	-4.44 *	-4.44 *	-1.96
	(0.73)	(0.82)	(0.82)	(1.02)
Connectedness$_{t-1}$	1.82 *	1.78 *	1.78 *	2.24 *
	(0.36)	(0.41)	(0.41)	(0.46)
YearsSinceLastEvent	-1.04 *	-1.54 *	-1.54 *	-1.08 *
	(0.07)	(0.14)	(0.14)	(0.13)
$Spline_1$	-0.00	-0.17 *	-0.17 *	-0.06 *
	(0.00)	(0.03)	(0.03)	(0.01)
$Spline_2$	-0.04 *	0.04 *	0.04 *	0.02 *
	(0.01)	(0.01)	(0.01)	(0.01)
$Spline_3$	0.01 *	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
$LogPopulation_{t-1}$	0.29 *	0.32 *	0.32 *	0.09
	(0.04)	(0.04)	(0.04)	(0.05)
$PopulationGrowth_{t-1}$	0.02	0.06	0.06	-0.08 *
	(0.02)	(0.03)	(0.03)	(0.03)
$LogGDPPerCapita_{t-1}$	0.02	-0.28 *	-0.28 *	-0.26 *
	(0.04)	(0.05)	(0.05)	(0.07)
$GDPPerCapitaGrowth_{t-1}$	-0.00	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
$Democracy_{t-1}$	-0.30 *	0.06	0.06	0.02
	(0.04)	(0.04)	(0.04)	(0.05)
$InternationalWar_{t-1}$	0.32	-0.33	-0.33	-0.30
	(0.20)	(0.21)	(0.21)	(0.24)
$CivilWar_{t-1}$ war	0.71 *	1.32 *	1.32 *	1.73 *
	(0.13)	(0.14)	(0.14)	(0.16)
$MilitaryRegime_{t-1}$	-0.12	-0.11	-0.11	-0.17
	(0.12)	(0.14)	(0.14)	(0.18)
BritishColonialLegacy	0.05	-0.32 *	-0.32 *	-0.25
	(0.11)	(0.14)	(0.14)	(0.17)
N	3829	3829	3829	3829
$\log L$	-1230.35	-959.03	-959.03	-619.10

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 2.6: Parameter estimates for logistic regressions of selected covariates on extreme violations of political imprisonment, torture, extrajudicial killing and disappearances.

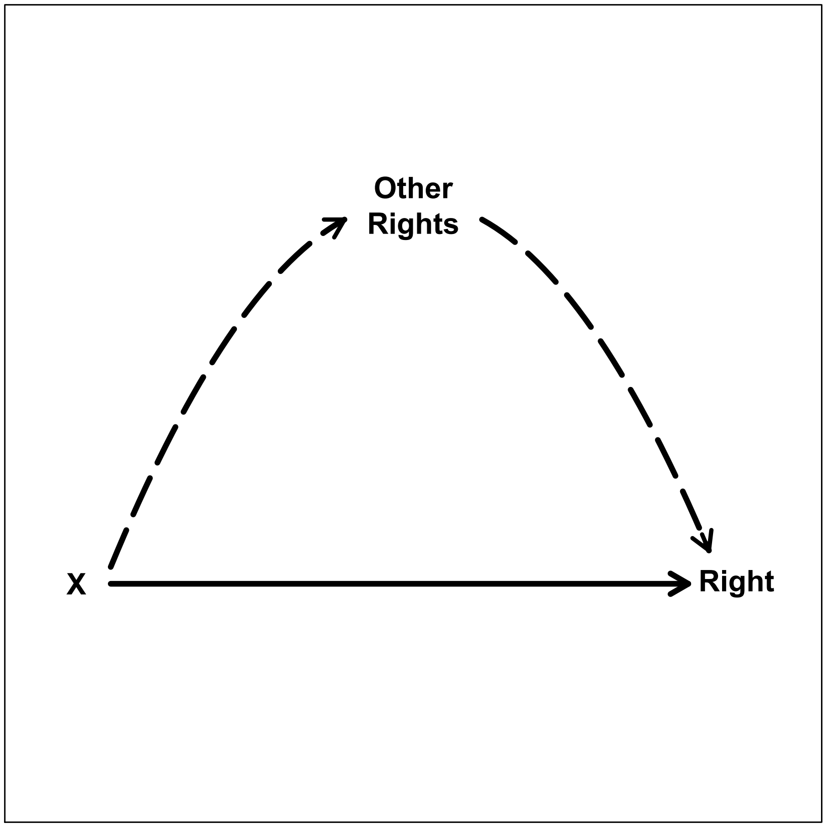


Figure 2.1: The causal variable *X* may affect both the specific human right under investigation as well as other human rights, which in turn may affect the specific human right. We conceptualize *X* as a *cost* or *constraint*. The network variables developed in this paper provide a way to model the interdependent relationships captured by this diagram.

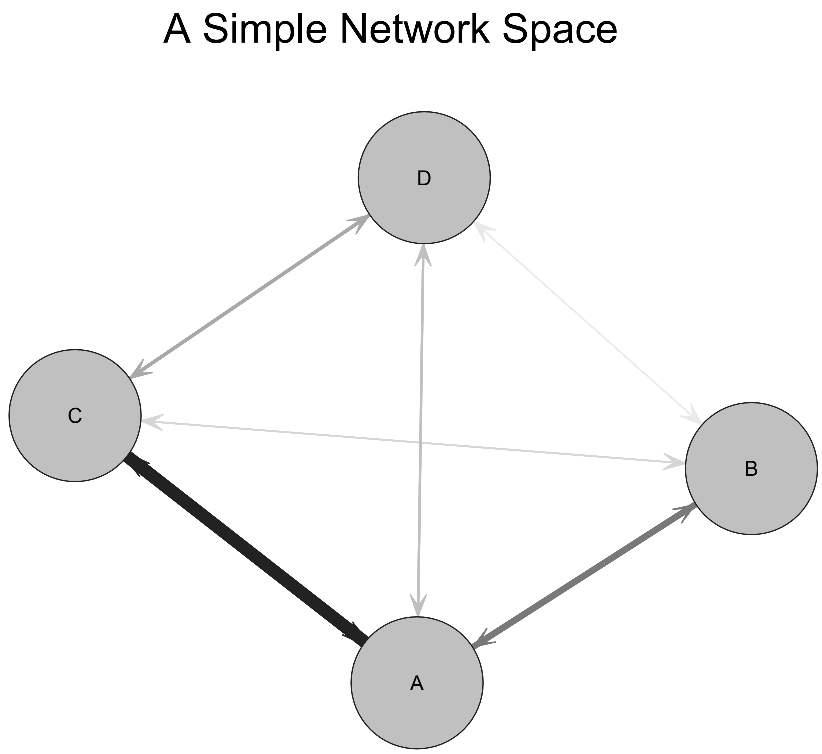


Figure 2.2: The proximity values (edges) link the four rights (nodes *A*, *B*, *C* and *D*) within the network. The weight and shade of the edges correspond to the proximity value; thus, the largest, darkest edge between right A and right C represents the largest proximity value of 0.9 while the thinnest and lightest edge between right B and right D represents the smallest proximity value of 0.1. Some values in this network are symmetric while others are not. For example, the proximity value that links right C to right D and the proximity value that links right D to right C are equivalent, while the proximity value that links right A to C and proximity value that links right C to A are asymmetric. The arrows indicated the direction of the proximity relationship such that $i \leftarrow j = P(i=1|j=1) - P(i=1|j=0)$. The arrows do not represent causal paths.

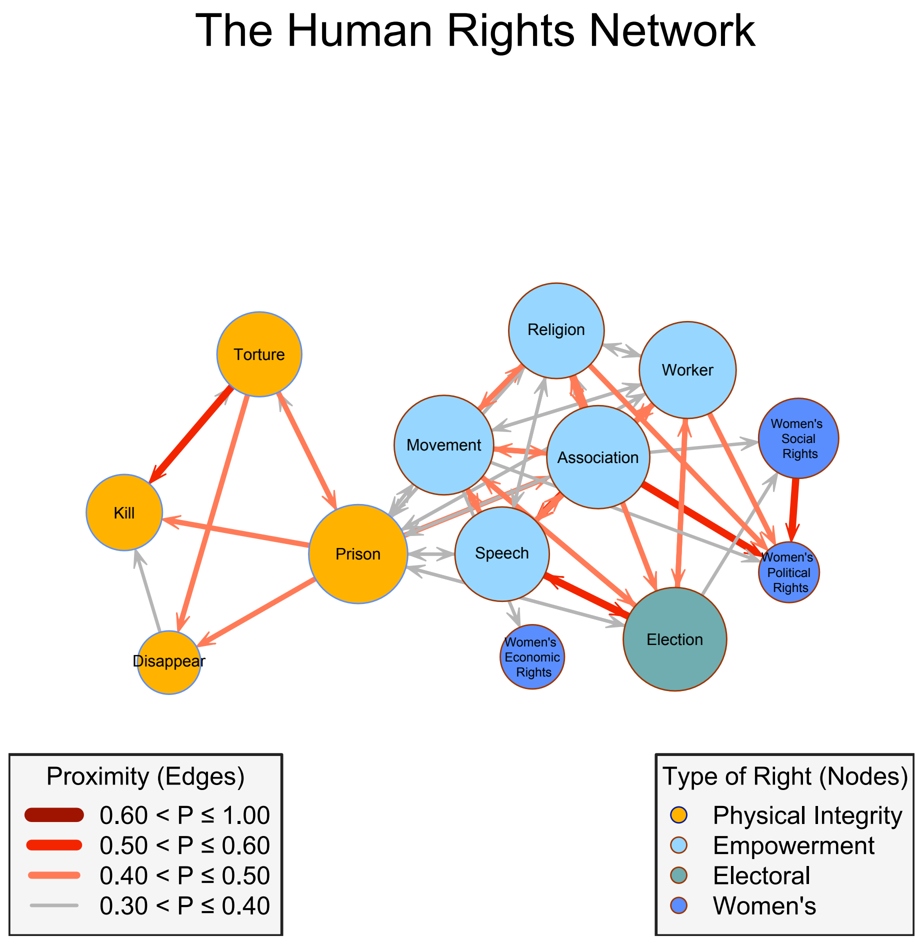


Figure 2.3: The human rights network, with human rights as nodes and proximity values ϕ_{ij} as edges. The plot is generated for all $\phi_{i,j,t} > 0.3$ between extreme violations in the average year. The node sizes are proportional to $\sum_j \phi_{i,j,t}$ and represent the influence of one right on all other rights in the network. The arrows should be interpreted for $i \leftarrow j$ as $P(i = 1|j = 1) - P(i = 1|j = 0)$. The arrows do not represent causal paths.

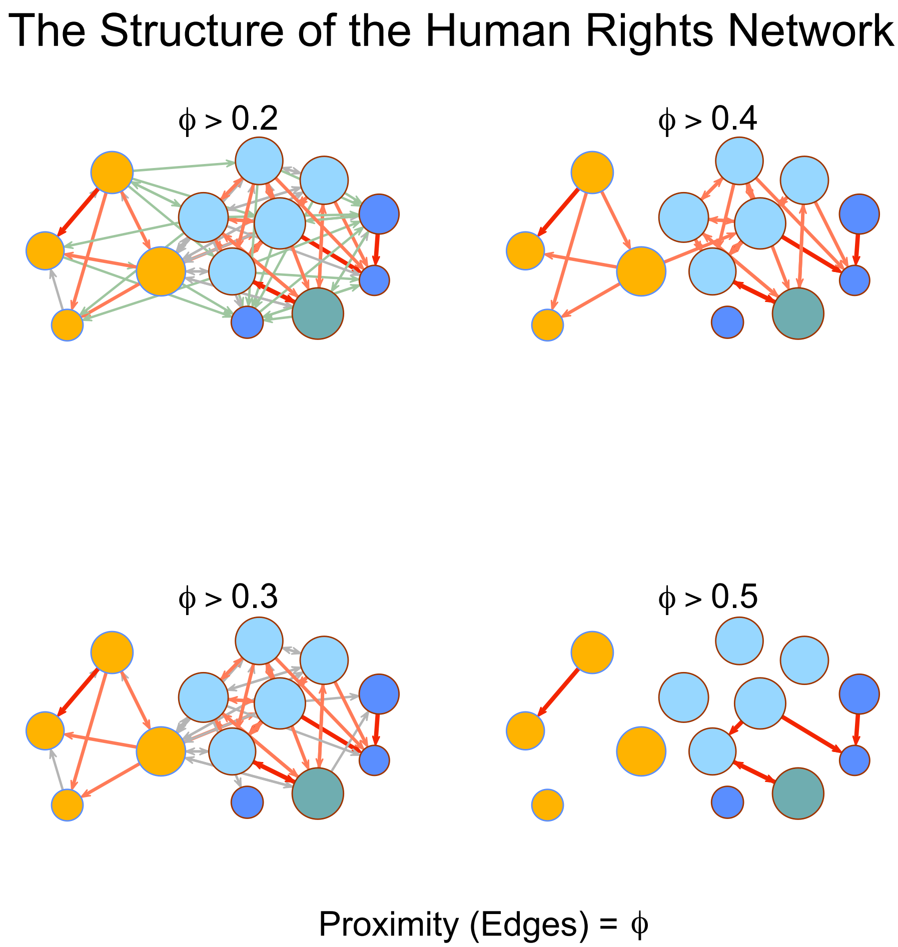


Figure 2.4: The four plots are generated with several proximity $\phi_{i,j,t}$ values to reveal some of the dominant linkages within the the human rights network. The placement of the human rights nodes is identical to those in the network displayed in Figure 2.3.

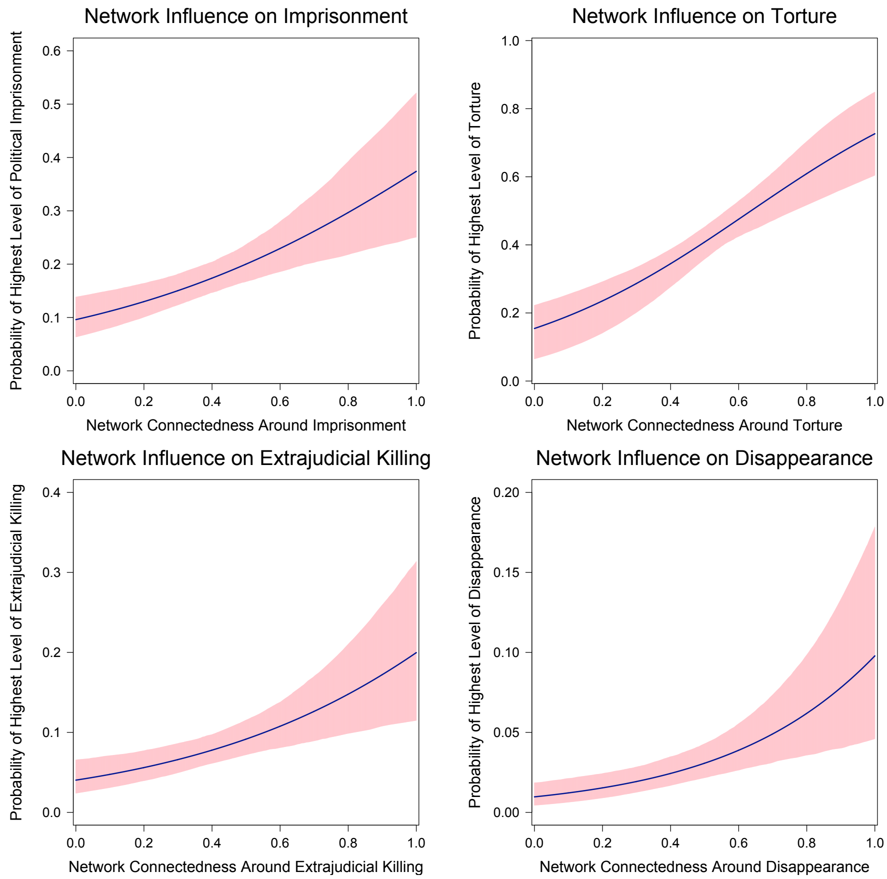


Figure 2.5: The expected value and 99% confidence intervals for the probability of extreme violations of the right over the range possible values of connectedness.

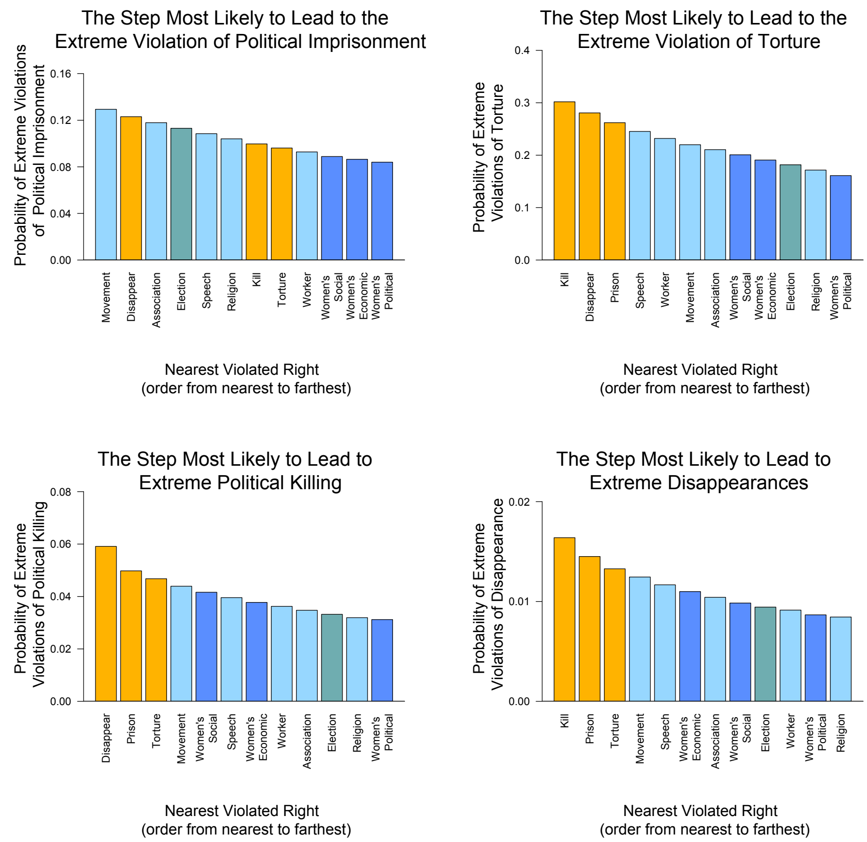


Figure 2.6: The probability of extreme violations of a right given the "nearest" violated right in the previous year. The x-axis is ordered by the proximity score. The information in this figure statistically confirms the pattern observed in Figure 2.3 and demonstrates the step most likely to lead to the extreme violation of the four physical integrity rights.

2.6 Appendix

2.6.1 Substitutes

Table 2.7: Substitution of High Level Repressive Action i for High Level Repressive Action j (1981-2006)

Action i	Action j	Year Count
Women's Political Rights	Torture	14
Torture	Women's Political Rights	14
Women's Political Rights	Extrajudicial Killing	11
Extrajudicial Killing	Women's Political Rights	11
Women's Political Rights	Disappearance	7
Disappearance	Women's Political Rights	7
Freedom of Religion	Disappearance	6
Disappearance	Freedom of Religion	6
Women's Social Rights	Disappearance	5
Freedom of Religion	Extrajudicial Killing	4
Women's Political Rights	Political Imprisonment	4
Political Imprisonment	Women's Political Rights	4
Disappearance	Women's Social Rights	4
Extrajudicial Killing	Freedom of Religion	4
Freedom of Assembly and Association	Disappearance	3
Disappearance	Freedom of Assembly and Association	3
Women's Social Rights	Extrajudicial Killing	2
Electoral Self-Determination	Disappearance	2
Women's Economic Rights	Disappearance	2
Disappearance	Electoral Self-Determination	2
Disappearance	Women's Economic Rights	2
Freedom of Movement	Women's Political Rights	2
Extrajudicial Killing	Women's Social Rights	2
Women's Political Rights	Freedom of Movement	2
Freedom of Movement	Disappearance	1
Worker's Rights	Disappearance	1
Women's Economic Rights	Torture	1
Women's Social Rights	Torture	1
Women's Economic Rights	Freedom of Assembly and Association	1
Women's Economic Rights	Freedom of Speech	1
Disappearance	Worker's Rights	1
Women's Social Rights	Worker's Rights	1
Freedom of Assembly and Association	Women's Economic Rights	1
Freedom of Speech	Women's Economic Rights	1
Torture	Women's Economic Rights	1
Torture	Women's Social Rights	1
Worker's Rights	Women's Social Rights	1
Disappearance	Freedom of Movement	1

2.6.2 Covariates

Descriptions for the variables used in the model presented in the main sections of this paper were taken from the Poe, Rost and Carey (2006) article. For theoretical justifications for these variables see the work by Poe and Tate (1994), Poe, Tate and Keith (1999) in addition to the short descriptions from the citations listed below.

- *Gross Domestic Product (GDP) per capita* is measured using the natural log of the country's gross domestic product in constant US dollars (1995) and reported per-capita. *GDP per capita growth* is measured as the yearly percentage change in GDP per capita. *Data Source*: World development indicators (World Bank 2009) and some missing values are taken from United States Energy Information Administration (2009). Since economic scarcity tends to increase tension and threats to the regime, nations with higher GDP and GDP growth are expected to be less likely to engage in repression (Poe and Tate 1994).

- *Population Size* is the natural log of a state's population estimate and *Population growth* is measured as the yearly percentage change in population. *Data Source*: World Development Indicators (World Bank 2009) and some missing values are taken from Fearon and Laitin (2003). Increased population and population growth are expected to be positively associated with repression, consistent with previous findings (Henderson 1991; Poe and Tate 1994; Poe, Tate and Keith 1999).

- *Level of Democracy* is measured using the Freedom House Political Rights scale. *Data Source*: Freedom House (2009) Poe, Rost and Carey (2006) reverse the scale of this variable. The result is a scale ranging from 1 (most democratic) to 7 (least democratic). Less democratic countries are expected to torture more frequently, so the effect of this variable is predicted to be positive (Henderson 1991; Poe and Tate 1994; Richards, Gelleny and Sacko 2001).

- *International War* This variable is coded 1 for participation in an interstate war or intervention in a civil war, 0 otherwise. *Data Source*: Uppsala Armed Conflict Dataset (Gleditsch et al. 2002; Harbom, Melander and Wallensteen 2008).

- *Civil War* This variable is coded 1 for civil war or intermediate conflict, 0 otherwise. *Data Source*: Uppsala Armed Conflict Dataset.(Gleditsch et al. 2002) The most recent update (Version 4-2008) to this data was conducted by Harbom, Melander and Wallensteen (2008).

- *Military Regime* This variable is coded 1 from the moment of a military coup until the military regime ceded government power, 0 otherwise. *Data Source*: Data are taken from several sources, including Madani (1992), *The Political Handbook of the World* (various years; see for example Banks and Muller, 1998) and the Central Intelligence Agency (CIA) (2004).

- *British colonial legacy* This variable is coded 1 if country was a British colony, 0 otherwise. *Data Source*: Poe, Tate and Keith (1999) and the Central Intelligence Agency (CIA) (2004).

2.7 Acknowledgements

This chapter is co-authored with Keith E. Schnakenberg and is forthcoming in the *Journal of Conflict Resolution*, which is a peer-reviewed political science journal. Earlier versions of this chapter were presented at the Political Networks Conference, June 11-13, 2009 and at the Annual Meeting of the American Political Science Association, September 3-6, 2009. For helpful comments, we thank Clair Apodaca, Robert Bond, Chad Clay, Courtenay Conrad, Christian Davenport, Jesse Driscoll, James Fowler, Jason Jones, Miles Kahler, David Lake, Yon Lupu, Dave Mason, Will Moore, Amanda Murdie, Jaime Settle, Reed Wood, Langche Zeng and the participants in the network analysis graduate seminar at UCSD. All data and code from this project are publicly available at a Dataverse archive maintained by the authors: `http://dvn.iq.harvard.edu/dvn/dv/HumanRightsScores` and are cross posted at the *Journal of Conflict Resolution* website: `http://jcr.sagepub.com/`. All questions regarding the data and code should be directed to the authors. Both authors contributed to study design, data collection, analysis, and preparation of the manuscript.

Chapter 3

Human Rights Treaty Compliance and The Changing Standard of Accountability

3.1 Introduction

The normative appeal of international law is predicated upon the view that well-designed rules will — in general and on average — promote peace, stability and good governance. (Goodman and Jinks, 2003: 171).

Why do states that ratify human rights treaties then violate human rights more often than those countries that do not ratify such treaties? Past research has puzzled over the empirical finding that countries that ratify the various instruments within the global human rights regime are more likely to abuse human rights than none ratifiers over time. In a recent paper, Fariss (2013) identified a possible answer to this question: the set of expectations used by monitoring agencies to hold states responsible for repressive actions has become increasingly stringent over time. Changes to this "standard of accountability" mask real improvements to the level of respect for human rights in data derived from monitoring reports, which implicitly incorporate these increasingly stringent assessments of state behaviors. Once this changing standard of accountability is taken into account using a new latent variable model of repression, the relationship between ratification of the UN Convention Against Torture and respect for human rights reverses and becomes positive. This result suggests that countries that respect human rights are more likely to ratify the UN Convention Against Torture in the first place, that the treaty has a causal effect on human rights protection once ratified, or possibly even both. This finding has broad implications for the international relations literature because it suggests that the human rights regime is not merely cover for human rights abusers and is in fact associated with respect for human rights and improvements in state behaviors over time. But what about the relationship between human rights respect and the many other human rights and humanitarian conventions that are apart of the global human rights regime?

In this paper, I demonstrate that there are systematic differences in the relationship between the level of respect for human rights across several widely studied human rights treaties. I demonstrate that this relationship holds more generally for a set of human rights treaties using a standard additive approach commonly used in the compliance literature to measure the level of embeddedness of a state within the international human rights regime over time. I also introduce a new measure of the global human rights regime using a dynamic latent variable model similar to the one developed by Martin and Quinn (2002). Overall, when changes in the standards used to assess state abuse are taken into account, the negative relationship between respect for human rights and ratification of human rights treaties is reversed. The results suggest that the "normative appeal of international law" may not be unfounded as earlier empirical research suggests.

In the remainder of this paper, I first define the standard of accountability and the mechanisms that cause changes to it over time. Second, I describe the disagreement that exists within the literature on international treaty compliance. The acknowledgment and assessment of the standard of accountability has major implications for this debate. Third, I describe the existing measurement strategies used to assess human rights behaviors and treaty compliance and then demonstrate how the new human rights data changes existing relationships with both old and new measurements of human rights treaty ratification over time. Finally, I conclude with suggestions for future research.

3.2 The Standard of Accountability and Treaty Compliance

The standard of accountability is the set of expectations that monitoring agencies use to hold states responsible for repressive actions (Fariss 2013). The standard of accountability has changed due to a combination of three mechanisms. These mechanisms influence the strategies and therefore the set of expectations that monitoring agencies use to assess and document state behaviors. First, improvements in the quality and increases in the quantity of information have led to more accurate assessments of the conditions in each country over time.[1] Second, access to countries by NGOs, like Amnesty International and Human Rights First (formerly the Lawyer's Committee for Human Rights), which seek to collect and disseminate accurate information about repression allegations and practices has increased as these organizations grow and cooperate with one another.[2] Third, changes in the subjective views of what constitute a "good" human rights record held by analysts at the monitoring agencies are anchored by the status quo, which improves as the global average of rights respect improves. Specifically, monitoring agencies are increasingly sensitive to the various kinds of ill-treatment that previously fell short of abuse but that still constitute violations of human rights.[3]

The set of expectations that monitoring agencies use to hold states responsible for repressive actions changes over time. The reports published today represent a broader and more detailed view of the human rights practices than reports published in previous years. As Sikkink notes, these monitoring agencies and others "have expanded their focus over time from a narrow concentration on direct government responsibility for the death, disappearance, and imprisonment of political opponents to a wider range of rights, including the right of people to be free from police brutality and the excessive use of lethal force" (2011, 159). Overall, the standard of accountability becomes more stringent as monitoring agencies like the US State Department and Amnesty International look harder for abuse, look in more places for abuse, and classify more acts as abuse.

[1] Keck and Sikkink (1998) attribute this change to an "information paradox". The paradox occurs when an increase in information leads to difficulties in assessing the efficacy of advocacy campaigns over time because of the very success in collecting and aggregating accounts of repressive actions in the first place. Clark and Sikkink (Forthcoming) coin a similar term — "human rights information paradox" — to describe this issue as it relates to human rights abuses specifically. As a result of this paradox, the global human rights situation may appear to have worsened over time because there is simply an increasing amount of information with which to assess human rights practices.

[2] Access to government documentation, witnesses, victims, prisons sites, and other areas are important for assessing state behaviors. Both Amnesty International and the US State Department rely on reports from other NGOs that collect and disseminate information about human rights abuses within states. The number and effectiveness of these actors has increased over time, especially since the end of the Cold War (Hopgood 2006; Hill Jr., Moore and Mukherjee Forthcoming; Korey 2001; Keck and Sikkink 1998; Lake and Wong 2009; Murdie and Bhasin 2011; Murdie and Davis 2012; Wong 2012).

[3] There is evidence from case law of a rising standard of acceptable treatment, whereby more acts come to be classified as inhuman treatment or torture. For example the European Court of Human Rights, in Selmouni v. France (1999), "consider certain acts which were classified in the past as Òinhuman and degrading treatmentÓ as opposed to ÒtortureÓ could be classified differently in future." That is, acts by state agents that might have previously been classified within the less severe category of ill-treatment and degrading punishment might now be classified as torture. The court states further "that the increasingly high standard being required in the area of the protection of human rights and fundamental liberties correspondingly and inevitably requires greater firmness in assessing breaches of the fundamental values of democratic societies." See Selmouni v. France, 25803/94, Council of Europe: European Court of Human Rights, 28 July 1999, available at: `http://www.unhcr.org/refworld/docid/3ae6b70210.html`.

What are the implications of the changing standard of accountability for state compliance with international human rights treaties? In the international relations literature there are two opposing viewpoints on treaty effectiveness. Authors such as Morrow (2007), Simmons (2000), and Simmons and Hopkins (2005) argue that treaty ratification constrains actors to modify their behaviors by creating costs for noncompliance. An alternative viewpoint is that countries only ratify a treaty if they would have complied even in the absence of the treaty. Thus, treaties have no effect on the behavior codified within the treaty such as the level of cooperation (e.g., Downs, Rocke and Barsoom 1996; Von Stein 2005), or ratification of certain human rights treaties (e.g., Hathaway 2002; Hafner-Burton and Tsutsui 2005, 2007; Hafner-Burton and Ron 2009). The theory presented here call into question this second viewpoint. Overall, the acknowledgment and assessment of the standard of accountability has major implications for this debate because it identifies changes to human rights reporting that correspond in time with the increasing embeddeness of countries within the global human rights regime.

In the remained of this paper I use the new human rights estimates presented by Fariss (2013), which are discussed in detail in that paper. The new estimates from the dynamic standard model provide strong evidence that physical integrity practices have improved over time. Unobserved changes in the standard of accountability explain why average levels of repression have, up until recently, appeared to remain unchanged as the constant standard model would suggest. Before proceeding with the tests, I discuss the estimation of a new variable, which measures the level of "embeddedness" of a country within the global human rights regime over time. I then compare the two human rights variables generated by Fariss (2013) with this new treaty variables, 2 additive scales, which are simply counts of the number of treaties a country has ratified, and 3 binary treaty variables. I describe all of these variables below.

3.3 Dynamic Latent Variable of Embeddedness Within the International Human Rights Regime

Examples of latent variable models are ubiquitous in political science, particularly in the study of legislative ideology in the United States (Poole 2005; Poole and Rosenthal 1991, 1997) but more recently in the study of comparative politics generally (Aleman and Saiegh 2007; Desposato 2006; Rosas 2009; Treier and Jackman 2008) and the study of human rights in particular (Schnakenberg and Fariss Forthcoming). Dynamic versions of these models are also increasingly common. The DW-NOMINATE procedure (Poole and Rosenthal 1997) is a dynamic version of W-NOMINATE, which estimates the ideal points of members of a legislature as a function of ideal points from the previous time period. Martin and Quinn (2002) introduced a Bayesian dynamic IRT model to estimate ideal points in the United States Supreme Court based on binary decision data. Martin and Quinn (2002) model the temporal dependence in the data by specifying a prior for each value of the latent variable centered at the estimated latent variable from the same unit in the previous time period. Schnakenberg and Fariss (Forthcoming) build on this insight in order to extend the ordinal IRT model introduced by Treier and Jackman (2008). More recently, Fariss (2013) extends this model further by allowing some of the item-difficulty parameters (the threshold parameters on the ordered variables) to vary over time instead of being held constant.

In this section of the paper, I use a dynamic binary IRT model, similar to the model introduced by Martin and Quinn (2002), to get a better sense of the level of “embeddedness” of a country within the global human rights regime over time. Below, I compare these new latent variable estimates to two alternative additive scales, which are simply the sum of the treaties a country has ratified in a given year.

The model is quite intuitive. The data are made up of country-year observations indexed by i and t. $i = 1,\ldots,N$ indexes countries and $t = 1,\ldots,T$ indexes year. Each of the $j = 1,\ldots,J$ indicators are for the various global treaties and optional protocols that cover a variety of human rights and humanitarian issues. Table 3.1 contains a listing of commonly studied human rights and humanitarian treaties that can be ratified by any country in the international system. The data can take on a value of “NA”, “0”, or “1”. A country-year observation is coded "1" for the year of ratification on treaty j and also 1 in every year following. The NAs are important because each of the treaties are available to sign at different points in time. Thus, treaties with more “missingness” will be less informative than those with less missingness over time. The uncertainty for each latent treaty variable for a given country year is based on the number of treaties that could potentially have been ratified in a given year. As Table 3.1 shows, fewer treaties exist in earlier decades. The latent variable therefore represents both the level of “embeddedness” of a country within the global human rights regime and the propensity to ratify new treaties as the open for signature over time. In a separate project, I am assessing the ratification behavior of states as a function of this new latent treaty variable.

Moreover, the model can also account for treaty reservations, two of which are considered in this paper. Accounting for treaty reservations is a difficult modeling problem. As Dancy and Sikkink (2012) state:

> It should be noted here that ratification is not a uniform process, and it may vary from state to state. For example, states may issue declarations of reservation upon ratification, which may express an unwillingness to accept in full the provisions of the treaty. However, we do not distinguish between types of ratification for two reasons: first, it is not commonly done in the quantitative literature, on which we are building; and second, it is quite difficult to generate a coding scheme that accounts for reservations. The reason is that any reduction in the score assigned to a reserving country would be arbitrary, given that reservations are not always similar. Future research should work to generate new ways of dealing with this issue (pg 771, note 51).

The dynamic binary IRT model can systematically account for such differences using a simply binary coding scheme. For this model, I have elected to include two binary variables, each coded 1 if a reservation does not exist for Article 21 and Article 22, conditional on ratification of the Convention Against Torture. If this convention is ratified but with a reservation to an article, then the binary indicators are coded 0. Otherwise the variable is coded as NA. Note again that the NA coding has substantive meaning in the context of the IRT model. Article 21 and Article 22 allow for the a special committee to review claims brought by other states (Article 21) and by individuals with the state (Article 22). It is not uncommon for states to ratify the Convention Against Torture with a reservation on this two Articles. Overall, the model provides a principled way to model this information along with the other data about treaty ratification generally.

To estimate the model, the error terms ε_{itj} are independently drawn from a logistic distribution, where $F(\cdot)$ denotes the logistic cumulative distribution function. Each of the j treaty variables are binary y_{itj} so the probability distribution is simply:

$$P[y_{itj}=1] = F(\alpha_j + \theta_{it}\beta_j) \tag{3.1}$$

Assuming local independence of responses across units, the likelihood function for β, α, and θ given the data is:

$$\mathscr{L}(\beta,\alpha,\theta|y) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[F(\alpha_j+\theta_{it}\beta_j)^{y_{itj}} * (1-F(\alpha_j+\theta_{it}\beta_j))^{(1-y_{itj})}\right] \tag{3.2}$$

Equation (1) refers to the probability of observing $y_{itj}=1$. 1 minus this equation refers to the probability of observing $y_{itj}=0$. The likelihood equation (2) refers to the probability of the observed value in the data y_{itj}. I estimate the model using independent standard normal priors on the latent treaty variable θ_{it}. In other words,

$$\theta_{it} \sim N(0,1)$$

for all i when $t=1$. The standard normal prior when $t>1$ is centered around the latent variable estimate from the previous year such that

$$\theta_{it} \sim N(\theta_{it-1},\sigma)$$

. This method for incorporating dynamics was implemented in the context of a dichotomous item-response theory by Martin and Quinn (2002). One difference between this model and the Martin and Quinn model is that σ is estimated instead of specifying it *a priori*.

The prior for variance σ is modeled as $U(0,1)$. This reflects prior knowledge that the between-country variation in will be much higher on average than the average within-country variance.[4] Slightly informative gamma priors Gamma$(4,3)$ were specified for the β parameters. The prior on β has strictly positive support to reflect the prior belief that all indicators contribute significantly and in the same direction to the latent variable. The α parameters are given $N(0,4)$ priors (extremely diffuse for this model), again for all of the j treaties.[5] The latent treaty model is estimated with two MCMC chains, which are

[4]The estimates of σ from the posterior of the converged model illustrate that the distribution is nowhere near 1, so the truncation decision was not important.

[5]As is generally true of item-response models, the likelihood function in is not identified. In particular, IRT models suffer from "invariance to reflection," which means that multiplying all of the parameters by -1 would have no effect on the likelihood function. Lack of identification is problematic in maximum likelihood models but is not a problem for Bayesian approaches. The problem of invariance to rotation motivated the choice to give the β parameters strictly positive priors. For more information on identification problems in IRT models, see Jackman (2009).

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).

The model assumes that any two item responses are independent conditional on the latent variable. This means that two item-responses are only related because of the fact that they are each an observable outcome of the same latent trait. In this case the latent trait is the level of embeddness of a country in the global human rights regime. There are three relevant local independence assumptions: (1) local independence of different indicators within the same country-year, (2) local independence of indicators across countries within years, and (3) local independence of indicators across years within countries. The third assumption is relaxed by incorporating temporal information into prior beliefs about the latent treaty variable.

I should reiterate that there is no model-free way to estimate a latent variable. An additive scale approach is a *model* assuming equally weighted indicators and no error. The new latent treaty variable provides an alternative to such a model by estimating the item-weights and the uncertainty of the estimates. I compare the latent variable, which is based on all of the treaty variables contained in Table 3.1 to two alternative scales. All of these variables are closely related. However, as Figure 3.1 illustrates, there is considerable overlap along the latent variable for cases that are assumed differences along the additive scales. The left panel compares the latent variable to an additive scale that ranges from 0 to 29 and is based on the total number of ratified treaties, optional protocols, and articles contained in Table 3.1. The right panel uses an alternative additive scale that ranges from 0 to 6 and is based on the total number of up to six ratified treaties. The selected treaties are commonly studied in the international relations literature and include the Convention Against Torture (CAT), the Covenant on Civil and Political Rights (CCPR), the Covenant on Economic, Social and Cultural Rights (CESCR), the Convention on the Elimination of All Forms of Racial Discrimination (CERD), the Convention on the Elimination of all Forms of Discrimination against Women (CEDAW), and the Convention on the Rights of the Child (CRC). To generate the additive scales, I recode missing values as 0 before then summing across the indicators. This is not necessary in the latent variable model because the model can account for missing data.

Not surprisingly, embeddeness within the international human rights regime has increased since the end of World War II. Figure 3.2 displays the average level of embeddedness over time. Figure 3.3 and Figure 3.4 display the rank order of countries by posterior mean in the year 1980 and 2000 respectively. The most embedded countries in 1980 are only as embedded as the middle ranked countries in the year 2000. In the next section I compare these treaty variables with the two human rights variables generated by Fariss (2013).

Table 3.1: Global International Human Rights Instruments

Treaty Name	**Signed**[1]	**Force**[2]
Convention on the Prevention and Punishment of the Crime of Genocide	1948	1951
Geneva Convention (1949)	1949	1949
International Convention on the Elimination of All Forms of Racial Discrimination	1965	1969
International Covenant on Civil and Political Rights	1966	1976
International Covenant on Civil and Political Rights Optional Protocol	1966	1976
International Covenant on Economic, Social and Cultural Rights	1966	1976
Convention on the Non-Applicability of Statutory Limitations to War Crimes and Crimes Against Humanity	1968	1970
International Convention on the Suppression and Punishment of the Crime of Apartheid	1971	1973
Geneva Convention Protocol I (relating to the Protection of Victims of International Armed Conflicts)	1977	1977
Geneva Convention Protocol II (relating to the Protection of Victims of Non-International Armed Conflicts)	1977	1977
Convention on the Elimination of all Forms of Discrimination against Women	1979	1981
Convention Against Torture	1984	1985
Convention Against Torture Article 21 (no reservation)	1984	1985
Convention Against Torture Article 22 (no reservation)	1984	1985
Convention on the Rights of the Child	1989	1990
International Covenant on Civil and Political Rights Optional Protocol	1989	1991
Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families	1990	2003
International Criminal Court	1998	2002
Convention on the Elimination of all Forms of Discrimination against Women Optional Protocol	1999	2000
Convention on the Rights of the Child Optional Protocol 1	2000	2002
Convention on the Rights of the Child Optional Protocol 2	2000	2002
Convention Against Torture Optional Protocol	2002	2006
Geneva Convention Protocol III (relating to the Adoption of an Additional Distinctive Emblem)	2005	2007
Convention on the Rights of Persons with Disabilities	2006	2007
Convention on the Rights of Persons with Disabilities Optional Protocol	2006	2008
International Convention for the Protection of All Persons from Enforced Disappearance	2006	2010
International Covenant on Economic, Social and Cultural Rights Optional Protocol	2008	2013
Convention on the Rights of the Child Optional Protocol 3	2011	——

Source 1: *University of Minnesota Human Rights Library* `http://www1.umn.edu/humanrts/`
Source 2: *United Nations Treaty Collections* `http://treaties.un.org/`
Note 1: The "Signed" column refers to the year the treaty is opened for signature.
Note 2: The "Force" column refers to the year the treaty enters into force.

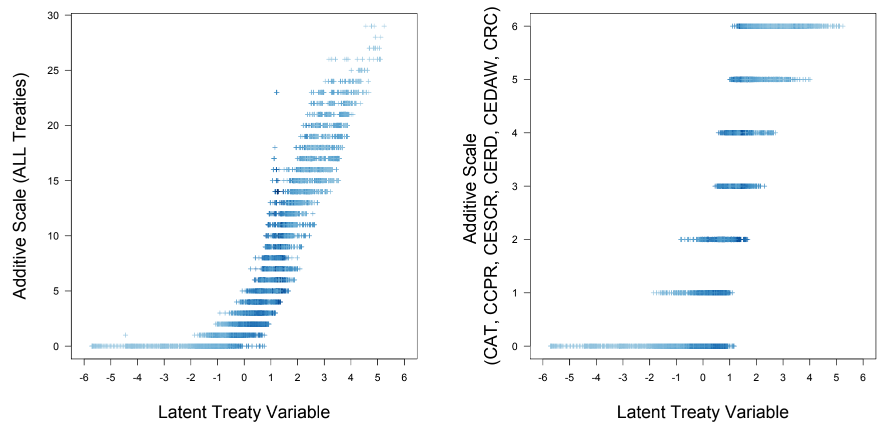


Figure 3.1: Relationship between the latent treaty variable and two additive scales. The left panel uses an additive scale that ranges from 0 to 29 and is based on the total number of ratified treaties and optional protocols contained in Table 3.1. The right panel uses an alternative additive scale that ranges from 0 to 6 and is based on the total number of up to six ratified treaties. The selected treaties are commonly studied in the international relations literature and include the Convention Against Torture (CAT), the Covenant on Civil and Political Rights (CCPR), the Covenant on Economic, Social and Cultural Rights (CESCR), the Convention on the Elimination of All Forms of Racial Discrimination (CERD), the Convention on the Elimination of all Forms of Discrimination against Women (CEDAW), and the Convention on the Rights of the Child (CRC). The correlation coefficients between the new latent variable and the additive scales are 0.856 $[95\% CI:\ 0.853,\ 0.8560]$ and 0.808 $[95\% CI:\ 0.804,\ 0.812]$ respectively. 95% credible intervals are generated by taking 1000 draws from the posterior distribution of latent treaty variable, which are then used to estimate the distribution of correlation coefficients.

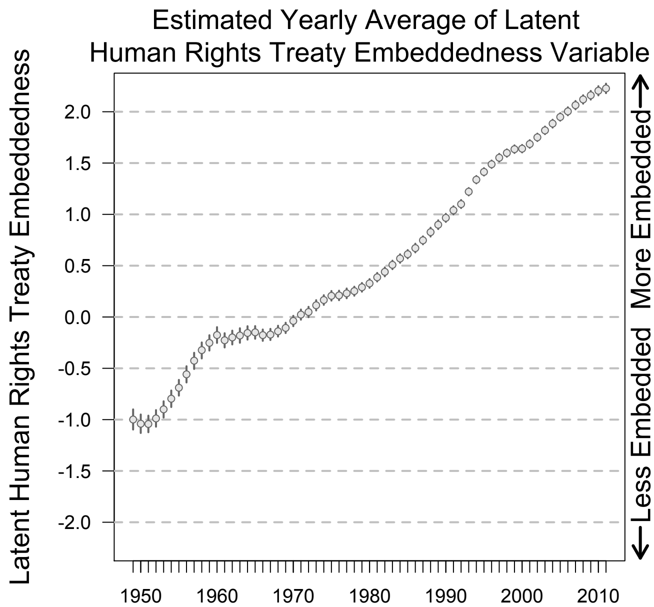
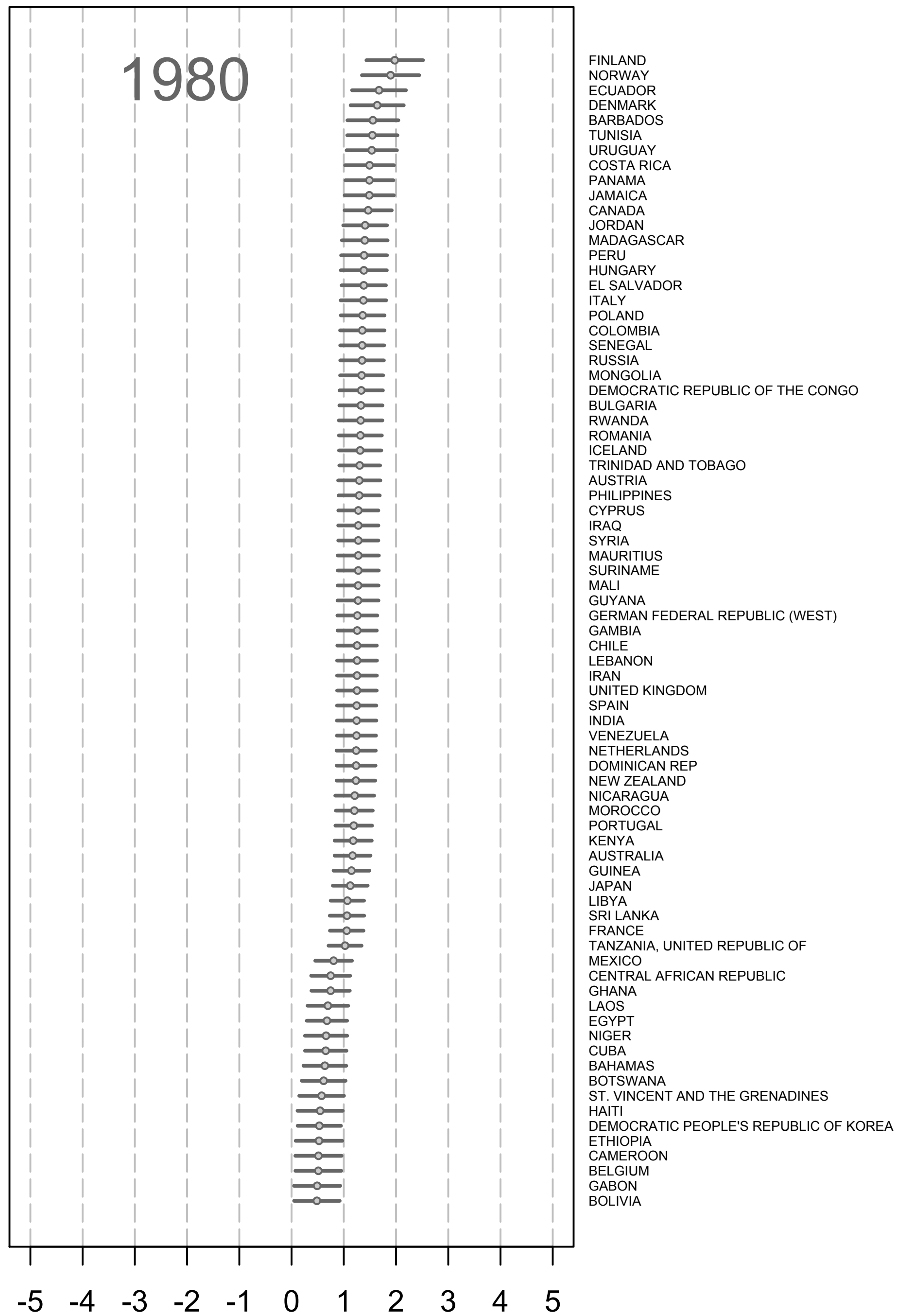


Figure 3.2: Average level of the human rights treaty variable over time. On average, countries become increasingly embedded in the global human rights regime over time.



1980
FINLAND
NORWAY
ECUADOR
DENMARK
BARBADOS
TUNISIA
URUGUAY
COSTA RICA
PANAMA
JAMAICA
CANADA
JORDAN
MADAGASCAR
PERU
HUNGARY
EL SALVADOR
ITALY
POLAND
COLOMBIA
SENEGAL
RUSSIA
MONGOLIA
DEMOCRATIC REPUBLIC OF THE CONGO
BULGARIA
RWANDA
ROMANIA
ICELAND
TRINIDAD AND TOBAGO
AUSTRIA
PHILIPPINES
CYPRUS
IRAQ
SYRIA
MAURITIUS
SURINAME
MALI
GUYANA
GERMAN FEDERAL REPUBLIC (WEST)
GAMBIA
CHILE
LEBANON
IRAN
UNITED KINGDOM
SPAIN
INDIA
VENEZUELA
NETHERLANDS
DOMINICAN REP
NEW ZEALAND
NICARAGUA
MOROCCO
PORTUGAL
KENYA
AUSTRALIA
GUINEA
JAPAN
LIBYA
SRI LANKA
FRANCE
TANZANIA, UNITED REPUBLIC OF
MEXICO
CENTRAL AFRICAN REPUBLIC
GHANA
LAOS
EGYPT
NIGER
CUBA
BAHAMAS
BOTSWANA
ST. VINCENT AND THE GRENADINES
HAITI
DEMOCRATIC PEOPLE'S REPUBLIC OF KOREA
ETHIOPIA
CAMEROON
BELGIUM
GABON
BOLIVIA
-5 -4 -3 -2 -1 0 1 2 3 4 5

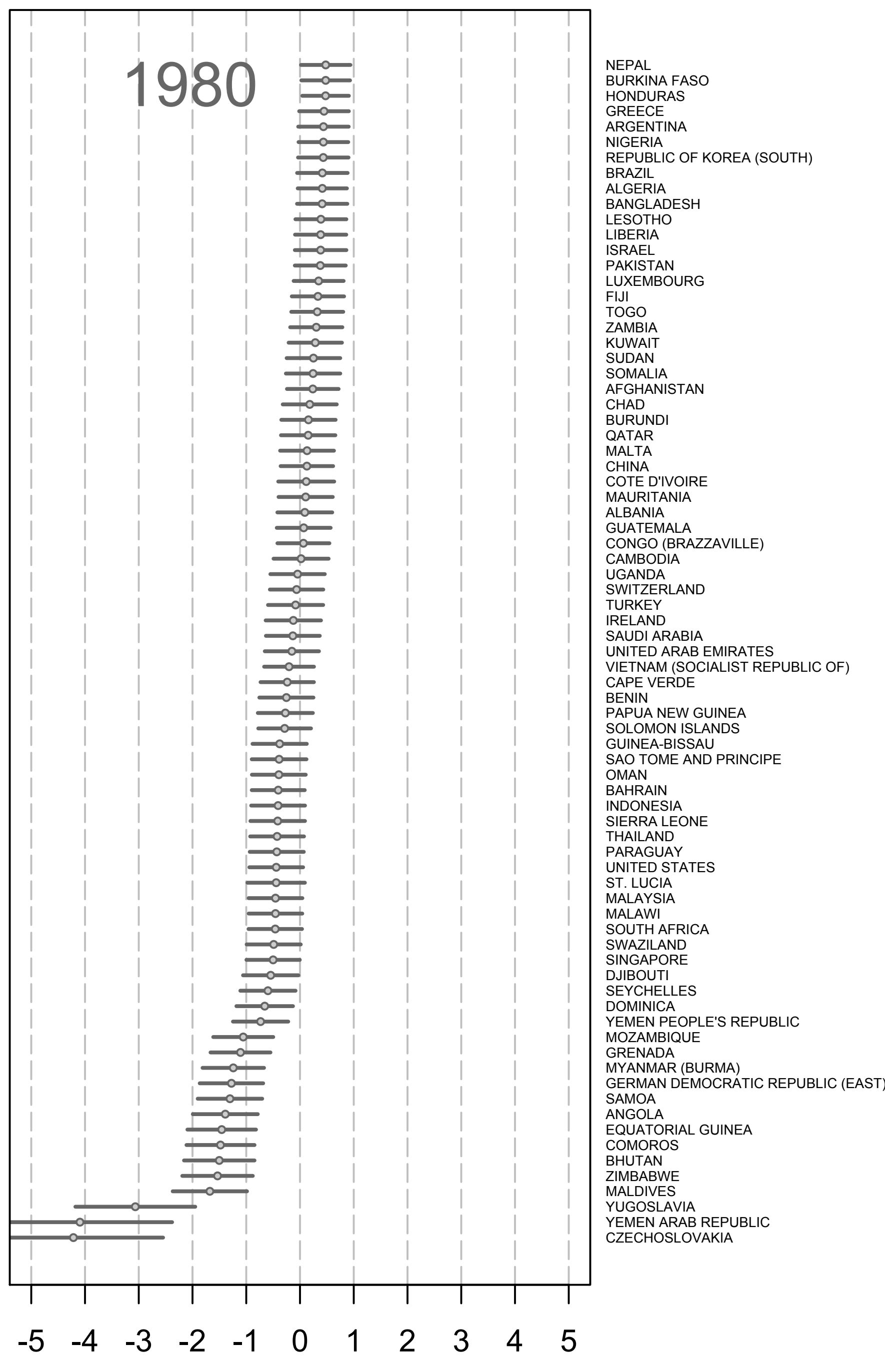
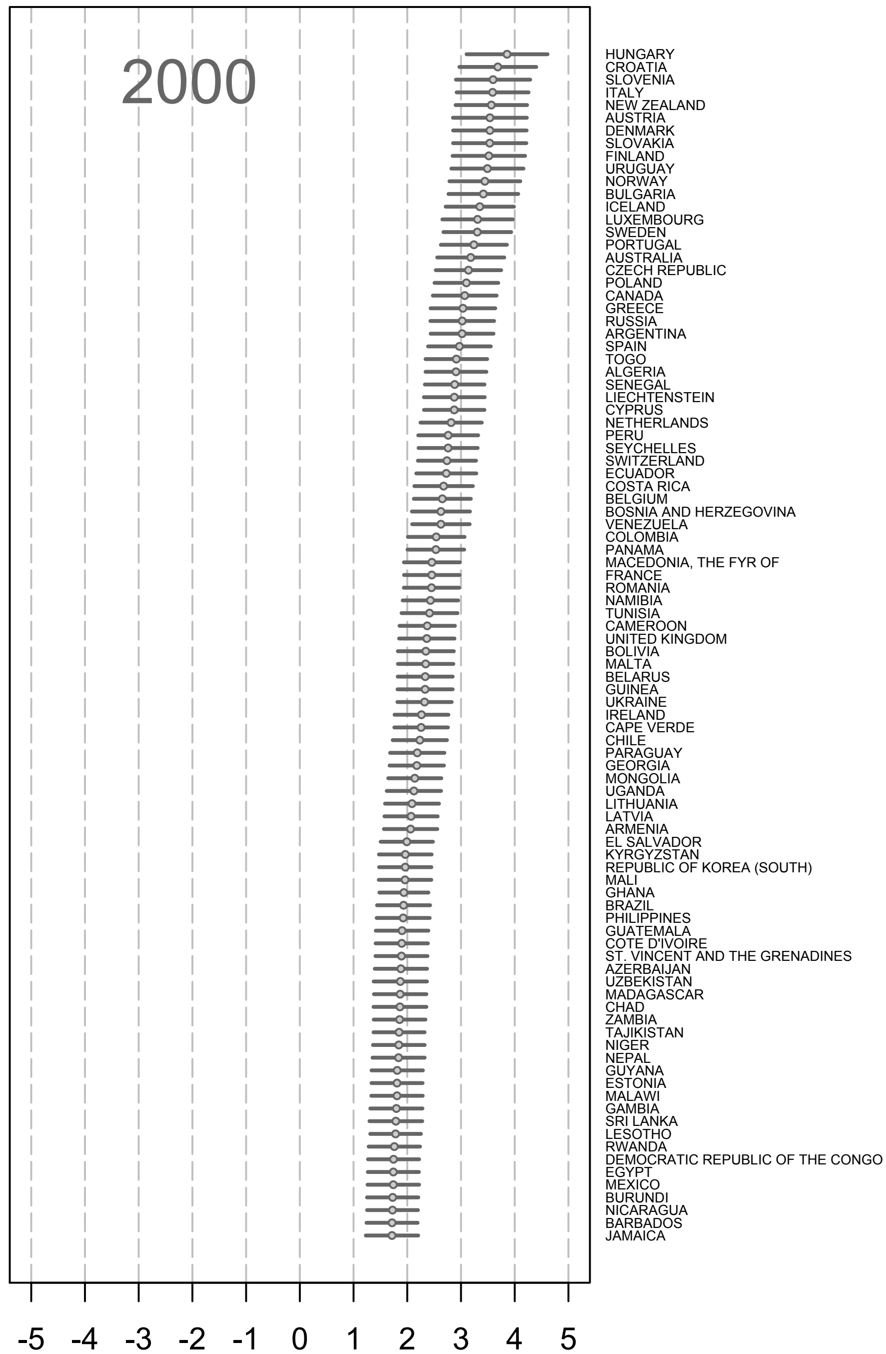


Figure 3.3: Rank order of countries by posterior mean in the year 1980. The most embedded countries in this year are only as embedded as the middle ranked countries in 2000.


2000
HUNGARY
CROATIA
SLOVENIA
ITALY
NEW ZEALAND
AUSTRIA
DENMARK
SLOVAKIA
FINLAND
URUGUAY
NORWAY
BULGARIA
ICELAND
LUXEMBOURG
SWEDEN
PORTUGAL
AUSTRALIA
CZECH REPUBLIC
POLAND
CANADA
GREECE
RUSSIA
ARGENTINA
SPAIN
TOGO
ALGERIA
SENEGAL
LIECHTENSTEIN
CYPRUS
NETHERLANDS
PERU
SEYCHELLES
SWITZERLAND
ECUADOR
COSTA RICA
BELGIUM
BOSNIA AND HERZEGOVINA
VENEZUELA
COLOMBIA
PANAMA
MACEDONIA, THE FYR OF
FRANCE
ROMANIA
NAMIBIA
TUNISIA
CAMEROON
UNITED KINGDOM
BOLIVIA
MALTA
BELARUS
GUINEA
UKRAINE
IRELAND
CAPE VERDE
CHILE
PARAGUAY
GEORGIA
MONGOLIA
UGANDA
LITHUANIA
LATVIA
ARMENIA
EL SALVADOR
KYRGYZSTAN
REPUBLIC OF KOREA (SOUTH)
MALI
GHANA
BRAZIL
PHILIPPINES
GUATEMALA
COTE D'IVOIRE
ST. VINCENT AND THE GRENADINES
AZERBAIJAN
UZBEKISTAN
MADAGASCAR
CHAD
ZAMBIA
TAJIKISTAN
NIGER
NEPAL
GUYANA
ESTONIA
MALAWI
GAMBIA
SRI LANKA
LESOTHO
RWANDA
DEMOCRATIC REPUBLIC OF THE CONGO
EGYPT
MEXICO
BURUNDI
NICARAGUA
BARBADOS
JAMAICA
-5
-4
-3
-2
-1
0
1
2
3
4
5

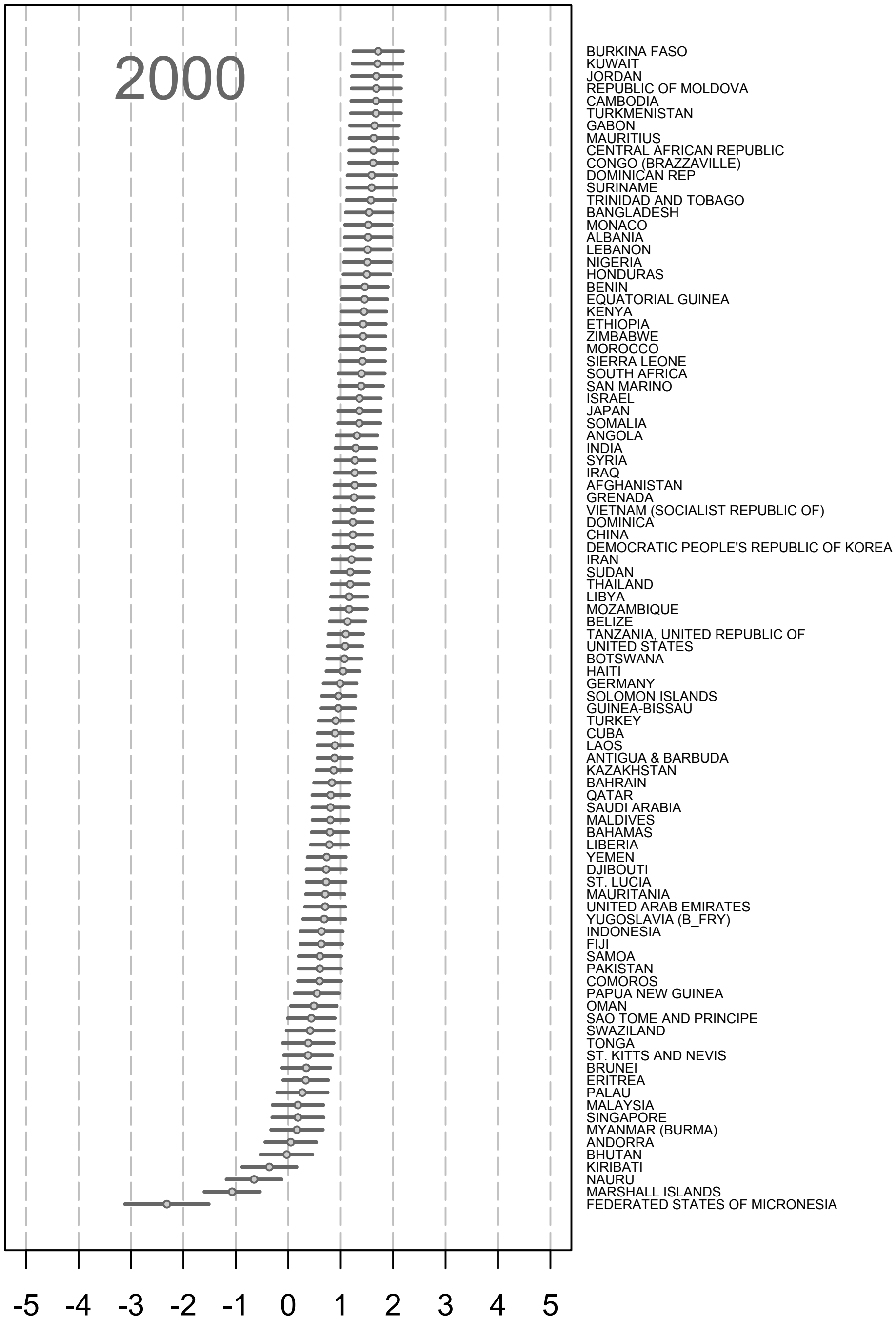


Figure 3.4: Rank order of countries by posterior mean in the year 2000.

3.4 Results: The Relationship Between Human Rights and Treaty Ratification

In this section, I illustrate the substantive importance of the changing standard of accountability for international relations theory by showing that ratification of UN human rights treaties and respect for physical integrity rights is positive. The result contradict negative findings from existing research. As the standard of accountability has increased over time, empirical associations with human rights data derived from standards-based documents and other variables will be biased if changes in the human rights documents are not accounted for. This is especially true for variables that measure the existence of institutions that are correlated with time such as whether or not the a particular treaty like the UN Convention Against Torture has been ratified or not.

I compare linear model coefficients using the dependent variable from the constant standard model and the dependent variable from the dynamic standard model. I estimate two linear regression equations using the latent physical integrity variables from the two measurement models. I regress these variables on several treaty variables, including the latent treaty variable defined above, two versions of an additive treaty scale and three binary variables that each measures whether or not a country has ratified the Convention Against Torture (CAT), the Convention on the Elimination of all Forms of Discrimination Against Women (CEDW), or the International Covenant on Civil and Political Rights (CCPR) in a given year. I also include several control variables in eight different specifications. I include a measure of democracy (Marshall, Jaggers and Gurr 2013), the natural log of GDP per capita (Gleditsch 2002), the natural log of population (Gleditsch 2002), and the lagged value of the latent human rights variable and finally the lagged value of one of the six different treaty variables. Each model always includes the lagged version of one of the two human rights variable and the lagged version of one of the six treaty variables.

The 8 linear regression models are specified as follows:

model 1 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1}$

model 2 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_3 * Polity2_{t-1}$

model 3 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_3 * Polity2_{t-1} + \beta_4 * ln(gdppc_{t-1})$

model 4 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_3 * Polity2_{t-1} + \beta_4 * ln(gdppc_{t-1}) + \beta_5 * ln(population_{t-1})$

model 5 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_4 * ln(gdppc_{t-1}) + \beta_5 * ln(population_{t-1})$

model 6 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_4 * ln(gdppc_{t-1})$

model 7 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_5 * ln(population_{t-1})$

model 8 $y_{it} \sim \beta_0 + \beta_1 * y_{i,t-1} + \beta_2 * treaty_{t-1} + \beta_3 * Polity2_{t-1} + \beta_5 * ln(population_{t-1})$

Where y_{it} is the country-year repression variable generated from either the dynamic standard model or the constant standard model presented by Fariss (2013). Each regression model is estimated 2 by 6 by 8 times in order to compare the 2 competing dependent variables, using each of the 6 different treaty variables within 8 different model specifications. All models include the same set of country-year observations from 1976 through 2005. The key piece of information to consider from the various models is the difference between the coefficient estimated for the treaty variable for the two competing dependent variables. Each of the 6 figures below visually displays the two coefficients estimates for all 8 models using one of the 6 treaty variables. The 9th panel in each figure displays the difference between these coefficients across the 8 model specifications. Notice that though the individual coefficients for the treaty variables change across the 8 models, the differences between the coefficients are consistent across all 8 specifications for all 6 treaty variables. By estimating all 8 of the model specifications for both dependent variables I am able to demonstrate that the estimated difference between the coefficients using the corrected human rights variable (dynamic standard model) compared to the uncorrected human rights variable (constant standard model) are consistent and statistically different from 0. Thus, even though the individual coefficients change depending on the model specification, the differences are consistent, which is a substantively important finding that eliminates concern that the use of a particular control variable is driving the results. The differences between coefficients are therefore robust to variable selection.

New relationships between treaty ratification and the level of human rights respect are obtained by replacing the dependent variable derived from the constant standard model with the one from the dynamic standard model. Recall that the dynamic standard model accounts for the changing standard of accountability whereas the constant standard model does not. Results are displayed across six figures. For example, Figure 3.5 plots the linear model coefficient for the latent rights treaty variable. Again, each model uses one of the two latent physical integrity dependent variables and various control variables.

Figure 3.6 and Figure 3.7 display results using the two additive treaty scales. Figure 3.8, Figure 3.9, and Figure 3.10 plot the coefficient for CAT, CEDAW, and CCPR ratification respectively.

The linear regressions using the dependent variable from the constant standard model tend to generate negative coefficients, which corroborates the empirical pattern described in earlier work. Comparison with the regression coefficient from the models using the dependent variable from the dynamic standard model is quite different however. The coefficients flip signs and many are statistically significant when compared with 0 and the alternative coefficient from the constant standard model. These results suggest that human rights protectors are more likely to ratify human rights treaties, that the treaties may in fact have some causal effect on human rights protection, or possibly both. The coefficient for the various treaty variables flip signs in nearly every model permutation presented across the six figures. These findings suggest that human rights treaties are not merely cover for human rights abusers. Table 3.2 summarizes the information contained in the 6 figures.[6] These findings call into question a key assumption about state behavior made by several recent papers about human rights treaty compliance (e.g., Hollyer and Rosendorff 2011; Vreeland 2008).

Table 3.2: Summary of results. Each figure display linear regression coefficients for one of two dependent variables regressed on the selected treaty variable and controls.

	Human Rights Latent Treaty Variable
Figure 3.5	Latent Human Rights Treaty Variable
Figure 3.6	Count of Selected Human Rights (CAT, CCPR, CESCR, CERD, CEDAW, CRC)
Figure 3.7	Count of All Human Rights (see Table 3.1)
Figure 3.8	Convention Against Torture (CAT)
Figure 3.9	Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW)
Figure 3.10	International Covenant on Civil and Political Rights (CCPR)

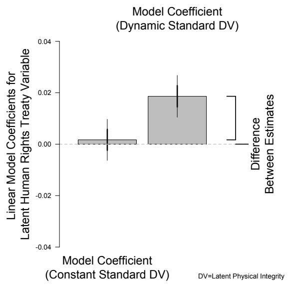
Note: Each treaty variable is included in each of 8 model specifications.

[6]Note that these models are not designed for causal inference and, though a variety of selection issues are known to exist when using this specification (see discussions in Neumayer (2005), Simmons and Hopkins (2005), Von Stein (2005), Simmons (2009), Hill Jr. (2010), and most recently Lupu (2013*b*)), the results from this type of model have spawned a large literature because of the counter intuitive, negative correlation found between ratification and respect for human rights (e.g., Hafner-Burton and Tsutsui 2005, 2007; Hathaway 2002; Hollyer and Rosendorff 2011; Vreeland 2008). Though this finding has been criticized (Clark and Sikkink Forthcoming; Goodman and Jinks 2003), it is generally taken for granted in the literature (Hafner-Burton and Ron 2009). Note however, that the selection issue is orthogonal to the differences in the two latent human rights variables used in the models.

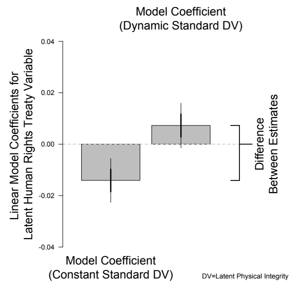
Human Rights Latent Treaty Variable

All Human Rights Treaties

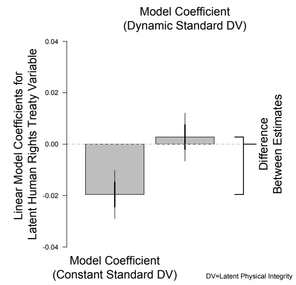
Model 1a



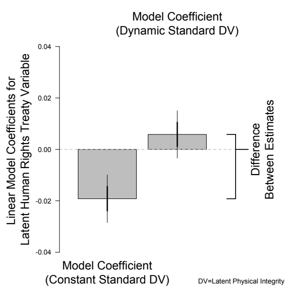
Model 2a



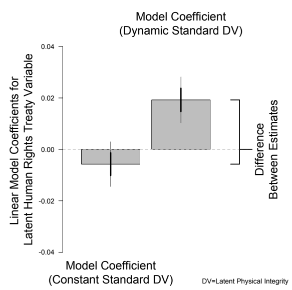
Model 3a



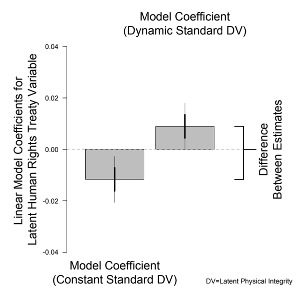
Model 4a



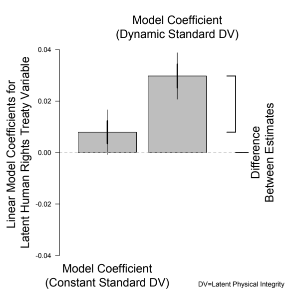
Model 5a



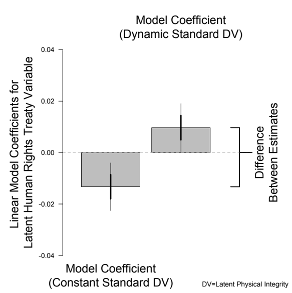
Model 6a



Model 7a



Model 8a



Treaty Coefficient Differences

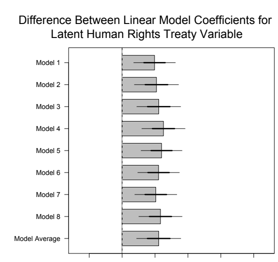


Figure 3.5: Estimated coefficient from the linear models using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error of the coefficient. The thin lines represent 2 $\pm$ the standard error of the coefficient. The differences are all consistent across models and all statistically different from 0.

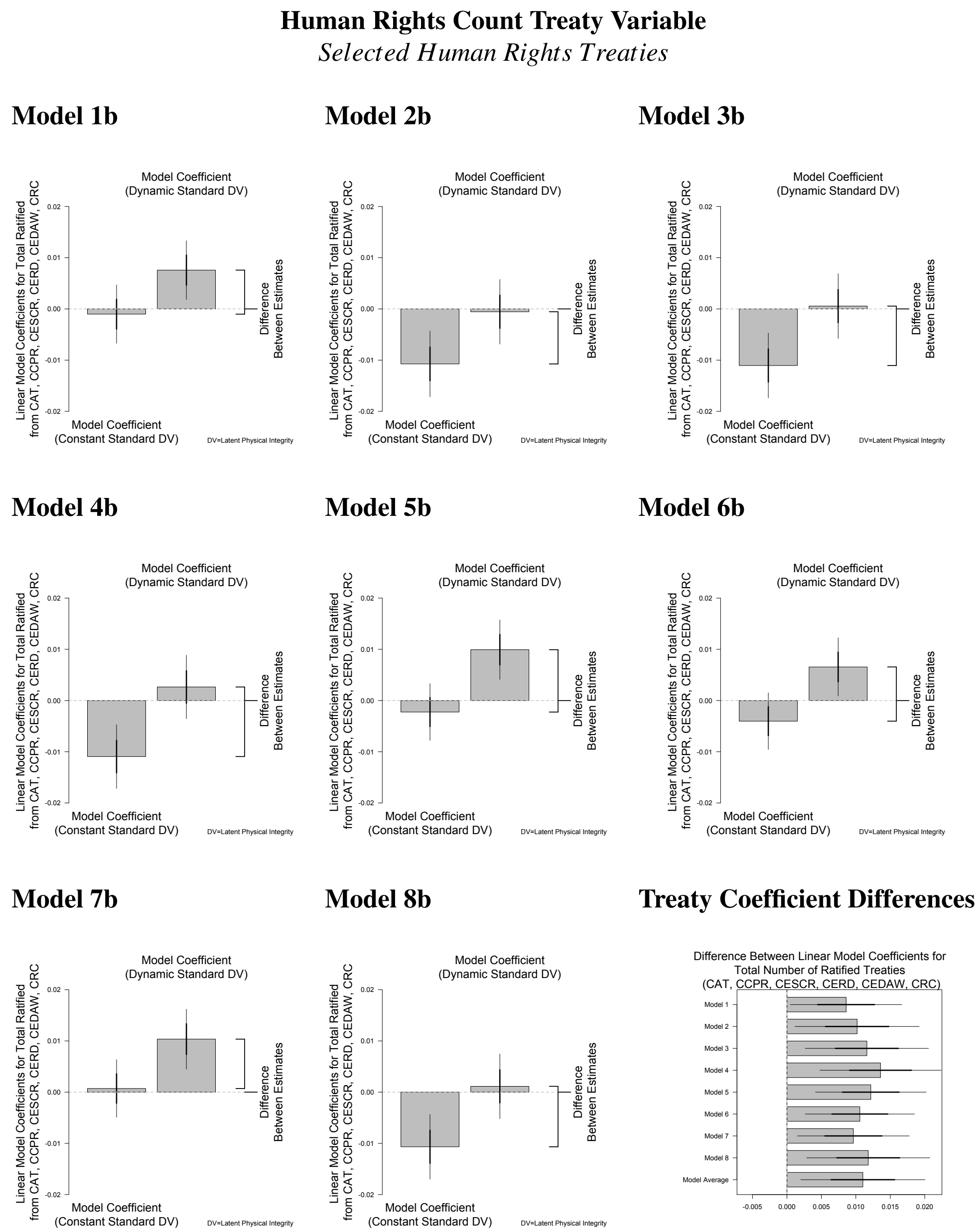
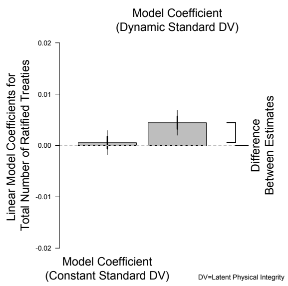


Figure 3.6: Estimated coefficient from the linear models using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error of the coefficient. The thin lines represent 2 $\pm$ the standard error of the coefficient. The differences are all consistent across models and all statistically different from 0.

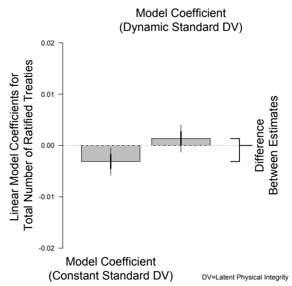
Human Rights Count Treaty Variable

All Human Rights Treaties

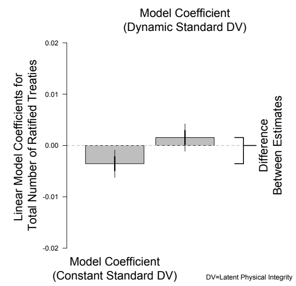
Model 1c



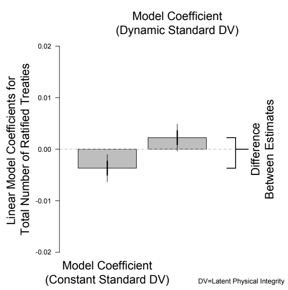
Model 2c



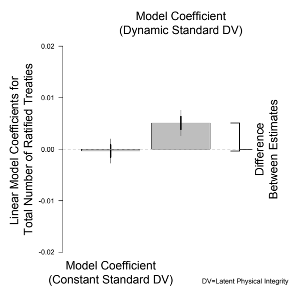
Model 3c



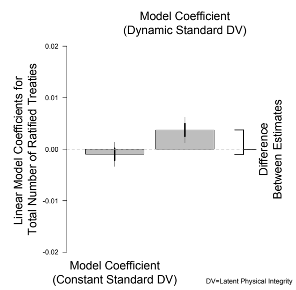
Model 4c



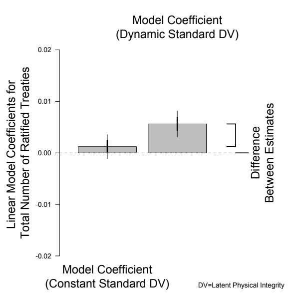
Model 5c



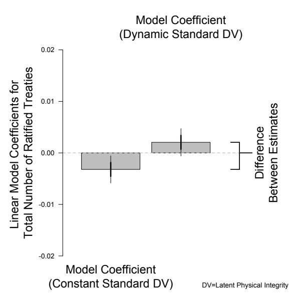
Model 6c



Model 7c



Model 8c



Treaty Coefficient Differences

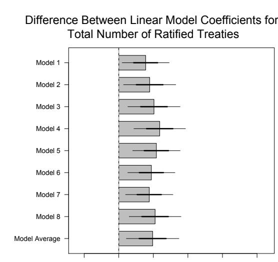


Figure 3.7: Estimated coefficient from the linear models using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error of the coefficient. The thin lines represent 2 $\pm$ the standard error of the coefficient. The differences are all consistent across models and all statistically different from 0. The average difference across the 8 models is 0.022 [95% CI : 0.008, 0.036].

Convention Against Torture

Binary Treaty Variable

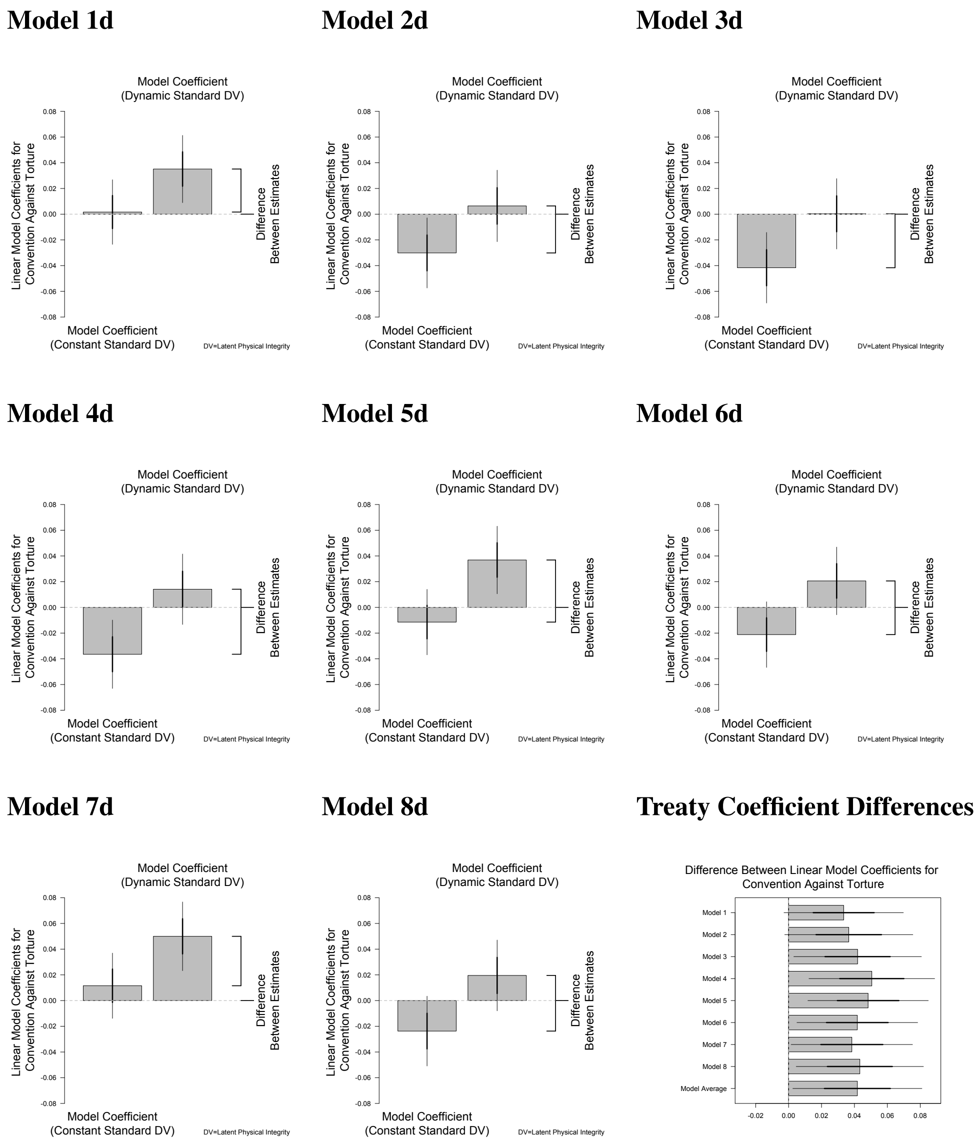


Figure 3.8: Estimated coefficient from the linear models using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 ± the standard error of the coefficient. The thin lines represent 2 ± the standard error of the coefficient. The differences are all consistent across models and all statistically different from 0.

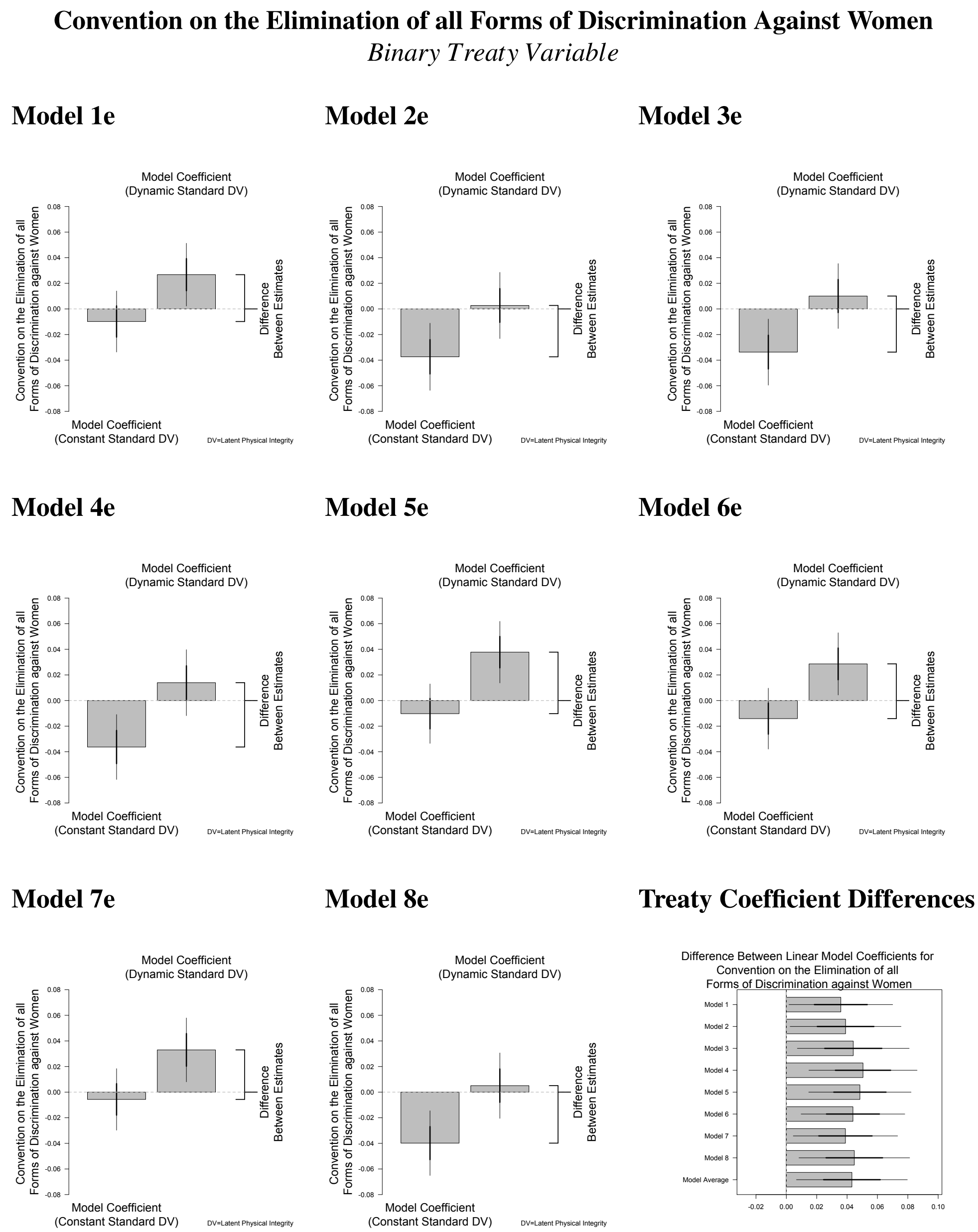


Figure 3.9: Estimated coefficient from the linear models using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 $\pm$ the standard error of the coefficient. The thin lines represent 2 $\pm$ the standard error of the coefficient. The differences are all consistent across models and all statistically different from 0.

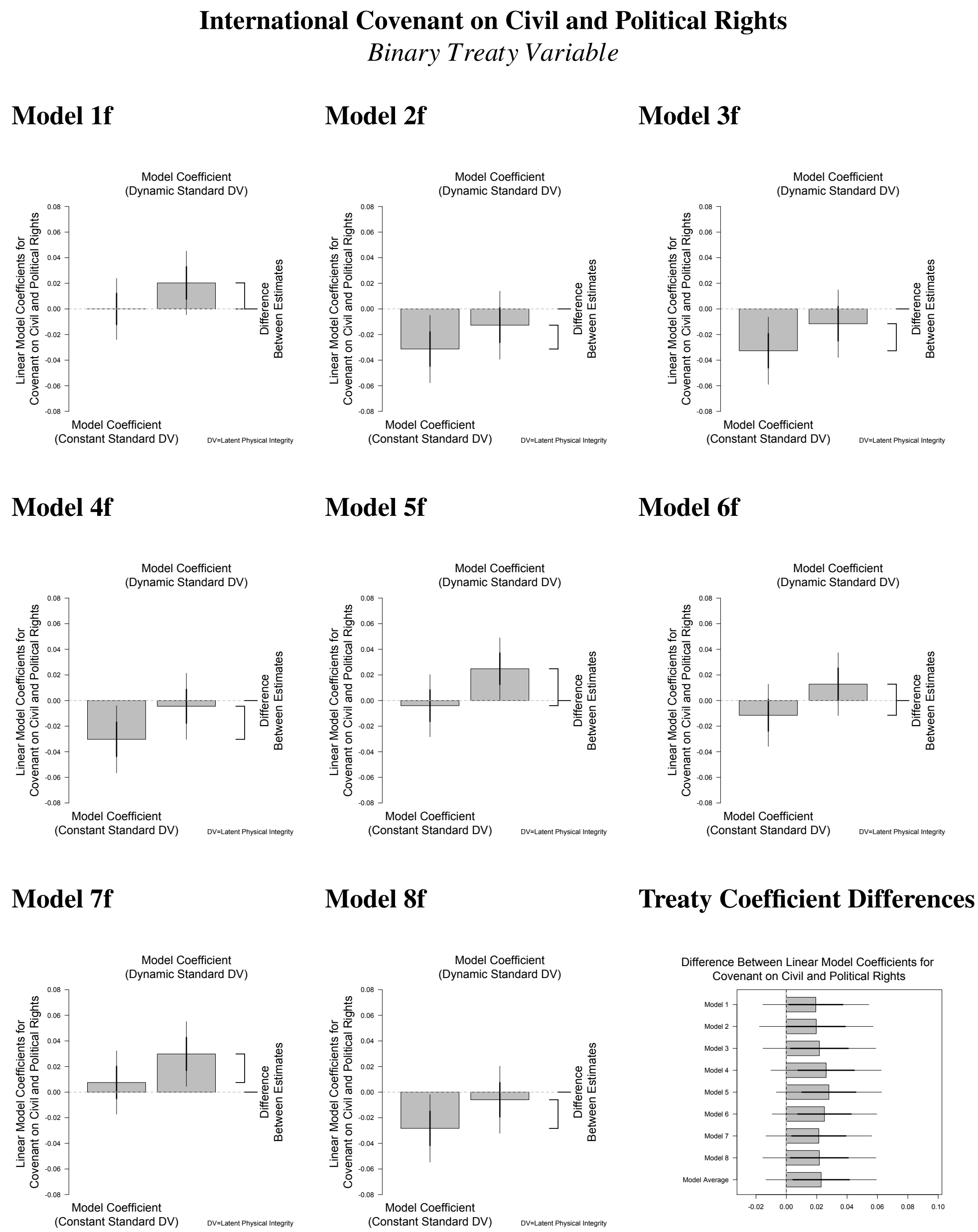


Figure 3.10: Estimated coefficient from the linear models using the dependent latent physical integrity variables from the constant standard model and the dynamic standard model respectively. The thick lines represent 1 ± the standard error of the coefficient. The thin lines represent 2 ± the standard error of the coefficient. The differences are all consistent across models and all statistically different from 0.

3.5 Conclusion

The results presented in this paper represent just the first step towards the reevaluation of what has become common knowledge in the literature of international treaty compliance. I have presented evidence that the ratification of human rights treaties is empirically associated with higher levels of respect for human rights over time and across countries. This evidence bolsters claims that the negative association between human rights abuse variables and treaty ratification is an artifact of some other un-accounted for process (Clark and Sikkink Forthcoming; Fariss 2013; Goodman and Jinks 2003). By accounting for the standard of accountability in new human rights data developed by Fariss (2013), a new picture has emerged of improving levels of respect for human rights, which coincides with the increasing embeddedness of countries within the international human rights regime. This positive relationship is robust to a variety of measurement strategies and model specifications. Much of the extant theorizing in international relations begins with the premies that international human rights treaties are not effective in order to explain this empirical anomaly (e.g., Hollyer and Rosendorff 2011; Vreeland 2008). The findings presented here cast considerable doubt on the empirical bases of such research.

Recent work by some human rights scholars has begun to unpack the institutional mechanisms by which international human rights treaties become effective (e.g., Dancy and Sikkink 2012; Kim and Sikkink 2010; Lupu 2013*a*; Sikkink 2011; Simmons 2009). Simmons (2009) looks specifically at the enforcement power of domestic institutions, whereas both Lupu (2013*a*) and Sikkink and her colleagues look at the intersection of domestic legal institutions and the specific human rights provisions within the various instruments that make up the international human rights treaty regime. Generally, this work considers the effectiveness of treaty compliance within countries that are democratic or at least in transition towards democracy.[7] The results presented here suggest that these authors should begin to reconsider the effectiveness of human rights treaties within non-democracies as well. With innovative new research designs available (e.g., Hill Jr. 2010; Lupu 2013*a*,*b*) and now new data and measurement tools as well (Fariss 2013; Schnakenberg and Fariss Forthcoming), scholars have the ability to begin the systematic reassessment of the role that international human rights treaties play in mitigating the use of repressive tactics by all types of governments.

[7]Lupu (2013*a*) analyzes all government types in his analysis of the role and effectiveness of domestic judicial systems in enforcing international treaty commitments.

3.6 Appendix

3.6.1 Regression Tables

Full regression results for the model differences presented in Figure 3.5 above are presented in Table 3.3 and Table 3.4.

Table 3.3: Linear Regression of Latent Human Rights (Dynamic Standard Latent Variable)

	Model 1a ^{Dynamic}		**Model 2a** ^{Dynamic}		**Model 3a** ^{Dynamic}		**Model 4a** ^{Dynamic}	
Variable	β (s.e.)	Z	β (s.e.)	Z	β (s.e.)	Z	β (s.e.)	Z
Intercept	-0.003 (0.007)	-0.388	0.007 (0.007)	0.988	-0.475 (0.053)	-9.002	-0.104 (0.063)	-1.642
Latent Human Rights$_{t-1}$	0.905 (0.005)	167.131	0.883 (0.006)	157.060	0.865 (0.006)	145.934	0.842 (0.006)	132.403
Latent Treaty$_{t-1}$	0.021 (0.005)	4.569	0.007 (0.005)	1.453	0.003 (0.005)	0.605	0.006 (0.005)	1.216
Polity2$_{t-1}$			0.010 (0.001)	9.698	0.007 (0.001)	7.467	0.010 (0.001)	9.807
ln(gdppc$_{t-1}$)					0.059 (0.006)	9.259	0.063 (0.006)	9.855
ln(Population$_{t-1}$)							-0.045 (0.004)	-10.390

	Model 5a ^{Dynamic}		**Model 6a** ^{Dynamic}		**Model 7a** ^{Dynamic}		**Model 8a** ^{Dynamic}	
Variable	β (s.e.)	Z	β (s.e.)	Z	β (s.e.)	Z	β (s.e.)	Z
Intercept	-0.325 (0.061)	-5.325	-0.583 (0.052)	-11.167	0.250 (0.039)	6.467	0.386 (0.041)	9.532
Latent Human Rights$_{t-1}$	0.862 (0.006)	139.713	0.877 (0.006)	147.350	0.895 (0.006)	157.219	0.862 (0.006)	141.472
Latent Treaty$_{t-1}$	0.017 (0.005)	3.643	0.012 (0.005)	2.568	0.025 (0.005)	5.454	0.010 (0.005)	2.087
Polity2$_{t-1}$							0.012 (0.001)	11.867
ln(gdppc$_{t-1}$)	0.078 (0.006)	12.285	0.072 (0.006)	11.268				
ln(Population$_{t-1}$)	-0.034 (0.004)	-8.092			-0.028 (0.004)	-6.675	-0.042 (0.004)	-9.569

Table 3.4: Linear Regression of Latent Human Rights (Constant Standard Latent Variable)

Variable	Model 1a^{Constant} β (s.e.)	Z	Model 2a^{Constant} β (s.e.)	Z	Model 3a^{Constant} β (s.e.)	Z	Model 4a^{Constant} β (s.e.)	Z
Intercept	-0.009 (0.007)	-1.298	0.001 (0.008)	0.163	-0.451 (0.053)	-8.461	-0.096 (0.064)	-1.509
Latent Human Rights$_{t-1}$	0.909 (0.005)	169.869	0.890 (0.005)	162.907	0.875 (0.006)	150.826	0.854 (0.006)	136.098
Latent Treaty$_{t-1}$	0.001 (0.005)	0.290	-0.014 (0.005)	-2.789	-0.019 (0.005)	-4.019	-0.019 (0.005)	-4.098
Polity2$_{t-1}$			0.009 (0.001)	8.788	0.007 (0.001)	6.485	0.009 (0.001)	8.715
ln(gdppc$_{t-1}$)					0.056 (0.006)	8.624	0.059 (0.006)	9.203
ln(Population$_{t-1}$)							-0.042 (0.004)	-9.873

Variable	Model 5a^{Constant} β (s.e.)	Z	Model 6a^{Constant} β (s.e.)	Z	Model 7a^{Constant} β (s.e.)	Z	Model 8a^{Constant} β (s.e.)	Z
Intercept	-0.302 (0.062)	-4.856	-0.553 (0.052)	-10.643	0.240 (0.040)	6.024	0.365 (0.040)	9.015
Latent Human Rights$_{t-1}$	0.870 (0.006)	144.714	0.884 (0.006)	155.201	0.899 (0.006)	162.744	0.871 (0.006)	149.209
Latent Treaty$_{t-1}$	-0.007 (0.005)	-1.591	-0.010 (0.005)	-2.243	0.004 (0.005)	0.947	-0.013 (0.005)	-2.812
Polity2$_{t-1}$							0.011 (0.001)	10.941
ln(gdppc$_{t-1}$)	0.073 (0.006)	11.492	0.067 (0.006)	10.558				
ln(Population$_{t-1}$)	-0.033 (0.004)	-7.797			-0.028 (0.004)	-6.367	-0.040 (0.004)	-9.172

3.6.2 Human Rights Data Sources

Table 4.5 and Table 4.6 contain information about the documentary sources used to generate each of the variables that enter the human rights latent variable models that generate the data used in this paper. For more information on these sources see the original citations and also Fariss (2013).

Table 3.5: Standards-Based Repression Data Sources

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
CIRI Physical Integrity Data, 1981-2010 - political imprisonment (ordered scale, 0-2) - torture (ordered scale, 0-2) - extrajudicial killing (ordered scale, 0-2) - disappearance (ordered scale, 0-2)	Cingranelli and Richards (1999, 2012*a*,*b*) Amnesty International Reports[1] and State Department Reports[2] *Information in Amnesty reports takes precedence over information in State Department reports*
Hathaway Torture Data, 1985-1999 - torture (ordered scale, 1-5)	Hathaway (2002) State Department Reports[1]
Ill-Treatment and Torture (ITT), 1995-2005 - torture (ordered scale, 0-5)	Conrad and Moore (2011), Conrad, Haglund and Moore (2012), Amnesty International (2006) Annual Reports[1], press releases[1], and Urgent Action Alerts[1]
PTS Political Terror Scale, 1976-2010 - Ammesty International scale (ordered scale, 1-5) - State Department scale (ordered scale, 1-5)	Gibney, Cornett and Wood (2012), Gibney and Dalton (1996) Amnesty International Reports[1] State Department Reports[1]

1. Primary Source; 2. Secondary Source

Table 3.6: Event-Based Repression Data Sources

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
Harff and Gurr Dataset, 1946-1988 - massive repressive events (1 if country-year experienced event 0 otherwise)	Harff and Gurr (1988) historical sources (see article references)[1]
Political Instability Task Force (PITF), 1956-2010 - genocide and politicide (1 if country-year experienced event 0 otherwise)	Harff (2003), Marshall, Gurr and Harff (2009) historical sources (see article references)[1] State Department Reports[2] Amnesty International Reports[2]
Rummel Dataset, 1949-1987 - genocide and democide (1 if country-year experienced event 0 otherwise) (3 death count estimates: best, low, high)	Rummel (1994*b*, 1995), Wayman and Tago (2010) New York Times[1], New International Yearbook[2], Facts on File[2], Britannica Book of the Year[2], Deadline Data on World Affairs[2], Kessing's Contemporary Archives[2]
UCDP One-sided Violence Dataset, 1989-2010 - government killing (event count estimate) (1 if country-year experienced event 0 otherwise) (3 death count estimates: best, low, high)	Eck and Hultman (2007), Sundberg (2009) Reuters News[1], BBC World Monitoring[1] Agence France Presse[1], Xinhua News Agency[1], Dow Jones International News[1], UN Reports[2], Amnesty International Reports[2], Human Rights Watch Reports[2], local level NGO reports (not listed)[2]
World Handbook of Political and Social Indicators WHPSI, 1948-1982 - political executions (event count estimate) (1 if country-year experienced event 0 otherwise)	Taylor and Jodice (1983) New York Times[1], Middle East Journal[2], Asian Recorder[2], Archiv der Genenwart[2] African Diary[2], Current Digest of Soviet Press[2]

1. Primary Source; 2. Secondary Source

3.7 Acknowledgements

I would like to thank Geoff Dancy, Jesse Driscoll, James Fowler, Miles Kahler, David Lake, Keith Schnakenberg, and Brice Semmens for many helpful comments and suggestions. The estimates from this paper along with the code necessary to implement the models in JAGS and R will be made publicly available at: `http://dvn.iq.harvard.edu/dvn/dv/HumanRightsScores`. This chapter is in preparation for submission to a peer-reviewed political science journal.

Chapter 4

A Dynamic Latent Variable Model of Human Rights Respect with Binary, Ordered, and Count Outcomes

4.1 Introduction

The counting of repressive events is difficult because state leaders have an incentive to conceal the actions of their subordinates and destroy evidence associated with abuse. In this paper I incorporate information on government killing from a variety of primary source documents into the measurement model of repression developed by Schnakenberg and Fariss (Forthcoming) and extended by Fariss (2013). By incorporating this information into the dynamic measurement model, I am able to generate new and more informative estimates of the number of specific repressive events and improve the estimation of the latent variable of repression presented by Fariss (2013).

Skepticism over the comparability of data that counts the number of repressive events in country-year observations was one of the main reasons for the movement away from event data in cross-national human rights research (Poe 2004). The model I describe below explicitly accounts for the uncertainty related to these counts and allows this uncertainty to help determine the degree of precision for each of the final country-year estimates produced by the model. The model estimates a latent variable of repression across space and time and includes both temporal information and information about missing values in the priors of the model.

In the remainder of this paper, I describe the existing latent variable model and then extend it to incorporate event data. I then introduce new count estimates generated from the model that also quantify the uncertainty inherent in the estimates of event-based count data by providing information about the underlying distribution from which such observations are drawn. Next, I demonstrate how the new latent variable estimates confirm the inferences generated in a recent analysis by Fariss (2013) and corroborate recent findings of a decline in the number of fatalities during war time (Goldstein 2011; Lacina, Gleditsch and Russett 2006) and a decline in the level of violence more generally (Pinker 2011). I also demonstrate how the new count estimates strengthen the inference from an existing study of UN peace keeping interventions (Hultman 2013) by accounting for the censoring of low levels of government killings. I close with a discussion of the promise of latent variable models for improving the measurement and study of repression and political violence generally.

4.2 Incorporating Event Count Data into the Latent Variable Model

Here I review the latent variable model parameterization developed by Schnakenberg and Fariss (Forthcoming) and extended by Fariss (2013). Formally, the statistical models compared by Fariss (2013) are both built on the assumption that the observed repression outcome variables for the country-year observations are each a function of the same underlying unidimensional latent variable, which represents the "true" or "latent" level of repression or respect for physical integrity rights. The goal of these and all other latent variables models is to estimate θ_{it}, which is the latent level of respect for physical integrity rights of country i in year t. The definition of "repression" or violations of "physical integrity rights" and sometimes called "state sanctioned terror" used here and in previous research includes arrests and political imprisonment, beatings and torture, extrajudicial executions, mass killing and disappearances, all of which are practices used by political authorities against those under their jurisdiction.[1]

For each model there are J indicators $j = 1,\ldots,J$. Some of the j indicators are ordinal with varying number of levels and some of the j indicators are binary. As already noted, $i = 1,\ldots,N$ indexes cross-sectional units and $t = 1,\ldots,T$ indexes time periods. y_{itj} is observed for each of the $j = 1,\ldots,J$ repression indicators. I let the values for each indicator be represented by k. In the original models, k is either ordinal or binary and can take on K_j values. For the binary indicators, $K_j = 2$. Below I extend the model to incorporate events counts, which can be any positive integer $k = 0,1,2,\ldots\infty$.

For each item, there is an "item discrimination" parameter β_j and a set of $K_j - 1$ "item difficulty cut-points" $\left(\alpha_{jk}\right)_{k=1}^{K_j}$. These parameters are analogous to a slope and intercept term in a logistic regression or the slope and cut-points in an ordered logistic regression. For the event count data there is again one "item discrimination" parameter β_j and a just one "item difficulty" parameter α_j for each of the observed event-based count variables.

The dynamic standard model parameterizes the difficulty cut-points for some of the items to vary over time such that $\left(\alpha_{tjk}\right)_{k=1}^{K_j}$. Note the t subscript here. This parameterization includes the standards-based variables from Cingranelli and Richards (1999, 2012*a*,*b*), Gibney, Cornett and Wood (2012), Hathaway (2002), and Conrad, Haglund and Moore (2012). The other items retain the constant item difficulty cut-point parameterization: $\left(\alpha_{jk}\right)_{k=1}^{K_j}$, which include the binary event-based variables drawn from Harff and Gurr (1988), Harff (2003), Rummel (1994*b*, 1995), Eck and Hultman (2007), Taylor and Jodice (1983).[2] Note the lack of a t subscript here. There is no t subscript on this parameter for any of the items in the constant standard model.

For the binary and ordered data, I assume error terms ε_{itj} are independently drawn from a logistic distribution, where $F(\cdot)$ denotes the logistic cumulative distribution function. The probability distribution for a given response to item j in the constant standard model is therefore given by

$$
\begin{aligned}
P[y_{itj} = 1] &= F(\alpha_{j1} - \theta_{it}\beta_j) && (4.1)\\
P[y_{itj} = k] &= F(\alpha_{jk} - \theta_{it}\beta_j) - F(\alpha_{jk-1} - \theta_{it}\beta_j) && (4.2)\\
P[y_{itj} = K] &= 1 - F(\alpha_{jK-1} - \theta_{it}\beta_j) && (4.3)
\end{aligned}
$$

For each item with constant difficulty cut-points, $y_{itj} = k$ if $\alpha_{jk-1} < \theta_{it}\beta_j + \varepsilon_{itj} < \alpha_{jk}$, and by specifying $\alpha_{j0} = -\infty$ and $\alpha_{jK_j} = \infty$ the probability equations (1), (2), and (3) reduce to[3]

[1] This definition is a modified version of one from Goldstein (1978).

[2] It is a coincidence that the event-based variables are each binary whereas the standards-based data are all categorical. The model is not dependent on this distinction.

[3] For each item with dynamic difficulty cutpoints, $y_{itj} = k$ if $\alpha_{tjk-1} < \theta_{it}\beta_j + \varepsilon_{itj} < \alpha_{tjk}$, where ε_{itj} is an error term and $\alpha_{tj0} = -\infty$ and $\alpha_{tjk_j} = \infty$.

$$P[y_{itj} = k] = F(\alpha_{jk} - \theta_{it}\beta_j) - F(\alpha_{jk-1} - \theta_{it}\beta_j) \tag{4.4}$$

Therefore, assuming local independence of responses across units, the constant standard's likelihood function for β, α, and θ given the data and model is $\mathscr{L}(\beta, \alpha, \theta | y, M_1)$ and is expressed as

$$\mathscr{L} = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j)\right] \tag{4.5}$$

The first set of equations (1), (2), and (3) and the reduced form (4) refer to the probability of observing a particular hypothetical level k. The likelihood equation (5) refers to the probability of the observed level in the data y_{itj}. These equations are the same for the dynamic standard model except for the addition of the t subscript on some of the α parameters. I have denoted this model as M_1 in order to distinguish between the extensions, which I describe next. Fariss (2013) compares at length differences between M_1, which he terms the constant standard model, and M_2, which he terms the dynamic standard model. I present M_2 next.

As a notational convenience let $v_j = 1$ when the j indicator is one of the standards-based variables and then $v_j = 0$ when it is one of the event-based variables. The probability distribution for the dynamic standard model is therefore

$$P[y_{itj} = k] = \left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j)\right]^{(v_j)} * \left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j)\right]^{(1-v_j)} \tag{4.6}$$

The likelihood function for the parameters given the data and model $\mathscr{L}(\beta, \alpha, \theta | y, M_2)$ is expressed as

$$\mathscr{L} = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[F(\alpha_{tjy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{tjy_{itj}-1} - \theta_{it}\beta_j)\right]^{(v_j)} * \left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j)\right]^{(1-v_j)} \tag{4.7}$$

Note that when $v_j = 0$, the probability distribution (6) and the likelihood function (7) for the dynamic standard model are equivalent to equation (4) and (5) for the constant standard model. The model is different when $v_j = 1$, which is when the standard of accountability changes over time.

4.3 Linking Event Count Data with a Latent Variable

The model can be further extended to take advantage of the event counts from some of the event-based data sources that were reduced to binary indicators in models M_1 and M_2. To make this extension, I simply use the negative binomial probability distribution to link the latent repression variable with the event count data. The systematic component for the repression event count variables is $exp(\alpha_j + \theta_{it}\beta_j)$, which is simply an alternative transformation of the linear component of the model parameters and the latent variable compared with the the logistic cumulative distribution function $F(\cdot)$. For the count repression variables α_j is the intercept, the item-difficult parameter, or the baseline expectation for the specific number of events. β_j is the slope coefficient or item-discrimination parameter. Finally, r_j is the over-dispersion parameter. Error terms ε_{itj} are independently drawn from a Gamma distribution in which the shape and rate parameters are equal.[4]

The negative binomial distribution arises from a variety of processes and can be parameterized in several ways. I have used the "ecological" parameterization of the negative binomial. The term "ecological" is not meant to imply that an ecological inference problem exists. It instead represents a count process arising from a system of heterogeneous units. The international system is clearly such an "ecological" system. Note that there is also a probabilistic parameterization for the negative binomial distribution, which is also known as the "failure-process" parameterization. These models are mathematically identical but are motivated by different phenomenological processes (see Bolker (2008), 165-167). JAGS only implements the probabilistic parameterization, so I re-parameterize the ecological model into the probabilistic one in the actual implementation of the model. The expected value of the probabilistic model in terms of the ecological model is $exp(\alpha_j + \theta_{it}\beta_j) = \frac{r_j*(1-p_{itj})}{p_{itj}}$ and the variance of the probabilistic parameterization in terms of the ecological parameterization is $exp(\alpha_j + \theta_{it}\beta_j) + \frac{exp(\alpha_j+\theta_{it}\beta_j)^2}{r_j} = \frac{r_j*(1-p_{itj})}{p_{itj}^2}$.

Note that the probabilistic parameterization assumes that r is a positive integer, whereas the ecological parameterization allows r to be a positive real number. This is useful for the statistical model, because it accounts for unobserved heterogeneity between units in the international system and not the number of successes in a set of trials. A smaller estimated value of r indicates an increasing amount of heterogeneity in the data. As r increases, the variance approaches the mean and the distribution therefore begins to approximate a Poisson distribution.[5]

---

[4]See Greene (2008) for a discussion. The error term can be expressed as either $exp(\alpha_j + \theta_{it}\beta_j + \varepsilon_{itj})$ or $exp(\alpha_j + \theta_{it}\beta_j) * exp(\varepsilon_{itj})$.

[5]The small size of the over-dispersion parameters indicates a high degree of heterogeneity in the data, which means the negative binomial is a good choice of estimator, relative to the Poisson. See King (1989) for a discussion of this choice when considering IR data. See Greene (2008) for additional discussion of this model and several alternatives. Greene (2008) also describes the relationship between the ε and r for the negative binomial model, also referred to as the NB2 model in the econometrics literature (e.g., Cameron and Trivedi 1986, 2005).

The probability distribution for a given response to item j in the event count data is therefore given by

$$P[y_{itj} = k] = \frac{\Gamma(r_j + k)}{\Gamma(r_j)k!} \left(\frac{r_j}{exp(\alpha_j + \theta_{it}\beta_j) + r_j} \right)^{r_j} \left(\frac{exp(\alpha_j + \theta_{it}\beta_j)}{exp(\alpha_j + \theta_{it}\beta_j) + r_j} \right)^{k} \tag{4.8}$$

The likelihood function for the parameters given the data and model $\mathscr{L}(\beta, \alpha, \theta, r|y, M_3)$ is expressed as

$$\mathscr{L} = \prod_{i=1}^{N} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\Gamma(r_j + y_{itj})}{\Gamma(r_j)y_{itj}!} \left(\frac{r_j}{exp(\alpha_j + \theta_{it}\beta_j) + r_j} \right)^{r_j} \left(\frac{exp(\alpha_j + \theta_{it}\beta_j)}{exp(\alpha_j + \theta_{it}\beta_j) + r_j} \right)^{y_{itj}} \right] \tag{4.9}$$

Equation (8) refers to the probability of observing a particular hypothetical level k. The likelihood equation (9) refers to the probability of the observed count in the data y_{itj}. M_3 is only useful for data that are all event counts. Next I combine M_2 and M_3 to create a model capable of handling ordered, binary, and count data.

Recall the notational convenience introduced above, where $v_j = 1$ when the j indicator is one of the standards-based variables and then $v_j = 0$ when it is one of the event-based variables. To incorporate the likelihood equation from M_2 and M_3, let $c_j = 0$ when the j indicator is binary or ordinal and then $c_j = 1$ when the j indicator is measured as a count.

Using the combination of v_j and c_j indicator variables allows for the combination of the likelihood equation (9) with the equation (7) together so that appropriate terms are reduced to 1. The probability distribution of the extended model is therefore:

$$\begin{aligned} P[y_{itj} = k] = & \left[\left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j) \right]^{(v_j)*(1-c_j)} \right] * \\ & \left[\left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j) \right]^{(1-v_j)*(1-c_j)} \right] * \\ & \left[\left[\frac{\Gamma(r_j + k)}{\Gamma(r_j)k!} \left(\frac{r_j}{exp(\alpha_j + \theta_{it}\beta_j) + r_j} \right)^{r_j} \left(\frac{exp(\alpha_j + \theta_{it}\beta_j)}{exp(\alpha_j + \theta_{it}\beta_j) + r_j} \right)^{k} \right]^{(1-v_j)*(c_j)} \right] \end{aligned} \tag{4.10}$$

The likelihood function for the parameters given the data and model $\mathscr{L}(\beta, \alpha, \theta, r|y, M_4)$ is expressed as

$$\mathscr{L} = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left[\left[F(\alpha_{tjy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{tjy_{itj}-1} - \theta_{it}\beta_j)\right]^{(v_j)*(1-c_j)}\right]*$$
$$\left[\left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j)\right]^{(1-v_j)*(1-c_j)}\right]*$$
$$\left[\left[\frac{\Gamma(r_j + y_{itj})}{\Gamma(r_j)y_{itj}!}\left(\frac{r_j}{exp(\alpha_j + \theta_{it}\beta_j) + r_j}\right)^{r_j}\left(\frac{exp(\alpha_j + \theta_{it}\beta_j)}{exp(\alpha_j + \theta_{it}\beta_j) + r_j}\right)^{y_{itj}}\right]^{(1-v_j)*(c_j)}\right] \quad (4.11)$$

To summarize M_4, equation (10) refers to the probability of observing a particular hypothetical level k and the likelihood equation (11) refers to the probability of the observed ordered, binary, or count data y_{itj}. Notice that for each of the j response variables that 2 of the 3 lines in equation (10) and (11) reduce to 1 so that only the appropriate link function contributes information to the likelihood. When $v_j = 1$ and $c_j = 0$ the first part of equation (11) contributes to the likelihood. When $v_j = 0$ and $c_j = 0$ the second part of equation (11) contributes to the likelihood. Finally, when $v_j = 0$ and $c_j = 1$ the third part of equation (11) contributes to the likelihood. In this way, different response variables can be used to help improve the estimation of the latent repression variable. In the next section I apply this model to the original 13 repression variables used by Fariss (2013) but replace the binary event-based from Eck and Hultman (2007) with the original count operationalization provided in this dataset (See Table 4.1 for the two operationalizations).

4.4 Application: One-Sided Government Killing

To validate this model, I focus on one dataset, which defines one-sided government killing as government caused deaths of non-combatants (Eck and Hultman 2007). The measurement of one sided government killing in which more than 25 individuals (non-combatants) are killed excludes extrajudical killings that occur inside a prison and combatant deaths that occur during civil conflicts (Eck and Hultman 2007). This data has several useful features that will help validate this model.

As Table 4.1 show, the UCDP data employ a variety of documentary sources in order to provide three estimates of one-sided government killing: $\{best, low, high\}$ (see the appendix for details about the coding). This data stems from the same underlying process, i.e., the latent level of repression, but it will be useful to model all three outcomes as function of the same underlying model with varying amounts of uncertainty for each of the three estimates.

The model introduced above (M_4) treats these three outcomes as distinct outcomes. That is, they are assumed to be independent. A useful analogy for this would be three different coders, one liberal (high), one conservative (low), and one moderate (best). The implementation of this model requires no alteration to the one presented above (M_4). An alternative and potentially more realistic parameteriza-

Table 4.1: One-sided government killing event-based repression data sources.

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
UCDP One-sided Violence Dataset, 1989-2010 - government killing (event count estimate) (1 if country-year experienced event 0 otherwise) (3 death count estimates: best, low, high)	Eck and Hultman (2007), Sundberg (2009) Reuters News[1], BBC World Monitoring[1] Agence France Presse[1], Xinhua News Agency[1], Dow Jones International News[1], UN Reports[2], Amnesty International Reports[2], Human Rights Watch Reports[2], local level NGO reports (not listed)[2]

1. Primary Source; 2. Secondary Source

tion of this model however, might consider one coder or set of coders deliberately making a distinction amongst three outcomes based on the available evidence in the various primary source documents. This model assumes that the α_J and β_J parameters are the same for the three one-sided government killing outcomes: $\{best,\ low,\ high\}$. The subscript on α_J and β_J is J in order to denote that these parameters are assumed to be the last value in the j vector of α and β parameters and therefore the same for each the three government-killing count variables or any additional event-based count data that is assumed to be an outcome of the same data generating process that produces them. This model is denoted as M_5. The only difference between the probability distribution (10) and likelihood (11) for M_4 and those for M_5 is the change from the j subscript to the J subscript for the α and β parameters in the negative binomial link portion of the equations.

The over-dispersion parameters r_j, which captures unobserved heterogeneity across units varies for each of the three outcome count variables. The estimation of a single item-difficulty parameter (intercept) and item-discrimination parameter (slope) for these three outcomes means that the variance of each country-year latent variable estimate $Var[\theta_{it}]$ will reduce (become more precise) when the three outcomes agree and will increase (become less precise) as the three outcomes diverge.

Conveniently, the quantity $exp(\alpha_J + \theta_{it}\beta_J)$ for the event-based count variables is the expected count of government killing. As should be clear, with the help of the model level parameters α_J, β_J, and the latent repression variable θ_{it}, I can estimate the expected number of government killings for all country-year observations in the data. Moreover, the uncertainty of this quantity is also quantifiable because the posterior distribution of each of the model parameters and the latent variable allows for the approximation of the posterior distribution of each country-year distribution of one-sided government killings. Increased uncertainty in θ_{it} leads directly to increased dispersion in the distribution of expected counts.

Model M_5 assumes one over-dispersion parameter r_j for each of the three one-sided government killing variables. However, it is simple to relax this assumption and generate country-year distributions of the number of individuals killed by their governments by estimating the same α_J and β_J parameters for the three one-sided government killing outcomes: $\{best,\ low,\ high\}$ and a unique over-dispersion parameter for each country year observation expressed as r_{it}. These distributions, by assumption, will always include the three original values reported by the producers of the one-sided government killing data. This is not the case for model M_5.

Model M_6 assumes a unique over-dispersion parameter r_{it} for each of the observations. Both of model $M)5$ and M_6 estimate each of the three count estimates $\{Best, Low, High\}$ as a function of the latent variable and a single item-difficulty parameter α_J and a single item-discrimination parameter β_J. The probability for model M_6 is

$$
\begin{aligned}
P[y_{itj} = k] = & \left[\left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j)\right]^{(v_j)*(1-c_j)}\right] * \\
& \left[\left[F(\alpha_{tjk} - \theta_{it}\beta_j) - F(\alpha_{tjk-1} - \theta_{it}\beta_j)\right]^{(1-v_j)*(1-c_j)}\right] * \\
& \left[\left[\frac{\Gamma(r_{it}+k)}{\Gamma(r_{it})k!}\left(\frac{r_{it}}{exp(\alpha_J + \theta_{it}\beta_J) + r_{it}}\right)^{r_{it}}\left(\frac{exp(\alpha_J + \theta_{it}\beta_J)}{exp(\alpha_J + \theta_{it}\beta_J) + r_{it}}\right)^{k}\right]^{(1-v_j)*(c_j)}\right]
\end{aligned}
\tag{4.12}
$$

The likelihood function for the parameters given the data and model $\mathscr{L}(\beta, \alpha, \theta, r|y, M_6)$ is expressed as

$$
\begin{aligned}
\mathscr{L} = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J} & \left[\left[F(\alpha_{tjy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{tjy_{itj}-1} - \theta_{it}\beta_j)\right]^{(v_j)*(1-c_j)}\right] * \\
& \left[\left[F(\alpha_{jy_{itj}} - \theta_{it}\beta_j) - F(\alpha_{jy_{itj}-1} - \theta_{it}\beta_j)\right]^{(1-v_j)*(1-c_j)}\right] * \\
& \left[\left[\frac{\Gamma(r_{it}+y_{itj})}{\Gamma(r_{it})y_{itj}!}\left(\frac{r_{it}}{exp(\alpha_J + \theta_{it}\beta_J) + r_{it}}\right)^{r_{it}}\left(\frac{exp(\alpha_J + \theta_{it}\beta_J)}{exp(\alpha_J + \theta_{it}\beta_J) + r_{it}}\right)^{y_{itj}}\right]^{(1-v_j)*(c_j)}\right]
\end{aligned}
\tag{4.13}
$$

Model M_6 is the preferred model because it relaxes the assumption that the level of one-sided government killing can be uniformly observed across units. Importantly, the country-year distributions for the set of count outcomes always includes the three original values reported by the producers of the one-sided government killing data, which is a useful feature for estimating the count data and then applying it. Relaxing this assumption comes at a cost however in terms of model complexity. Specifically, M_6 has about 10,000 more parameters to estimate compared to M_5. To determine if this alternative model is a better approximation of reality I conduct several tests of model fit. Nonetheless, this cost may be worth paying if precise estimates of the distribution of potential county-year one-sided government killings is the goal. Such data is quite useful in this application, which I demonstrate below.

The model parameters for the binary and ordered data are given the same prior distributions as in Fariss (2013) and displayed in Table 4.2 below. The additional parameters that link the latent variable with the count data introduced in this paper deserve further discussion. In model M_5, the over-dispersion parameter r_j is given a diffuse prior $U(0,100)$. For the model M_6, the over-dispersion parameters r_{it} are estimated for each country-year observation. To identify this more complex model, a hierarchical structure is imposed on the estimation of r_{it} using a random country intercept δ_i, and a random year intercept η_t. This structure actually helps to reduce the complexity of the model and speeds up estimation time. The intuition for this hierarchical parameterization is that information about the country and period of time will be informative for inferring the value of each country-year over-dispersion parameter r_{it} instead of assuming that they are independent. These parameters capture the level of heterogeneity or disagreement between each of the three count estimates as a function of the overall level of information in the country, which is captured by δ_i, and the overall level of information in the year, which is captured by η_t. The over-dispersion parameter is a deterministic function of these two random variables: $r_{it} = exp(\delta_i + \eta_t)$. As r_{it} approaches the expected count value, over-dispersion decreases and the relationship between the data reduces to a Poisson distribution. Small values of r_{it} therefore represent greater levels of heterogeneity in the data generating process.

To help identify the model I assume that the low estimate for government killing is 0 if missing from 1989-2010. I do not alter missing values for the best and high estimate. This decision increases the uncertainty of the resulting latent variable and decreases the over-dispersion parameters for both models (increases heterogeneity). Smaller over-dispersion parameters represent greater levels of heterogeneity in the underlying distribution. Importantly, countries with high levels of respect for human rights, as measured by the latent variable itself, can still have small estimated over-dispersion parameters but very low expected counts. The meaning of each over-dispersion parameter r_{it} is relative to the expected count.

The extended versions of the measurement model explicitly accounts for the uncertainty related to the estimated counts and allows this uncertainty to help determine the degree of precision for each of the final country-year latent variable estimates produced by the model. M_5 estimates the count specific heterogeneity as a global parameter for each of the three count variables. M_6 estimates count specific heterogeneity in the counts as a country-year parameter. Missing data does not lead to a loss of observations but only increases the uncertainty for a given estimate. Estimating this count data is not without its challenges however. Most importantly the number of primary sources available for each country varies and the quality and reliability of the information contained in each document varies as well. The model parameterizes each of these variables which will eventually allow researchers to make probabilistic statements about the relative quality of the information used in the estimation itself. This task will become increasingly important as more event-based count data is incorporated into the models developed here. I leave these tasks for future research.

The two new extended latent variable models M_5 and M_6 are implemented in R using Martyn Plummer's JAGS software (Plummer 2010) and compared below. Conventional diagnostics all suggested convergence including those of Geweke (1992), Heidelberger and Welch (1981, 1983), and Gelman and Rubin (1992), and standard graphical analysis.

Table 4.2: Summary of Prior Distributions for Latent Variable and Model Level Parameter Estimates

	M_5	M_6
Latent Variable		
country-year latent variable (first year)	$\theta_{i1} \sim N(0,1)$	$\theta_{i1} \sim N(0,1)$
country-year latent variable (other years)	$\theta_{it} \sim N(\theta_{it-1}, \sigma)$	$\theta_{it} \sim N(\theta_{it-1}, \sigma)$
uncertainty of latent variable	$\sigma \sim U(0,1)$	$\sigma \sim U(0,1)$
Model Parameters (Categorical Data)		
event-based variable cut-points (constant)	$\alpha_{jk} \sim N(0,4)$	$\alpha_{jk} \sim N(0,4)$
standards-based variable cut-points (first year)	$\alpha_{1jk} \sim N(0,4)$	$\alpha_{1jk} \sim N(0,4)$
standards-based variable cut-points (other years)	$\alpha_{tjk} \sim N(\alpha_{t-1,jk}, 4)$	$\alpha_{tjk} \sim N(\alpha_{t-1,jk}, 4)$
slope	$\beta_j \sim Gamma(4,3)$	$\beta_j \sim Gamma(4,3)$
Model Parameters (Count Data)		
event-based variable cut-points (constant)	$\alpha_J \sim N(0,4)$	$\alpha_J \sim N(0,4)$
slope	$\beta_J \sim Gamma(4,3)$	$\beta_J \sim Gamma(4,3)$
population over-dispersion rate	$r_j \sim U(0,100)$	
country-year over-dispersion rate		$r_{it} = exp(\delta_i + \eta_t)$
country-random effect for over-dispersion rate		$\delta_i \sim N(0,1)$
year-random effect for over-dispersion rate		$\eta_t \sim N(0,1)$

4.5 Results

In this section I first discuss the general predictive power of the two competing models (M_5 and M_6). Second, I demonstrate how the new latent variable estimates confirm the inferences generated in a recent analysis by comparing the latent variable from the model M_6 with those presented in this other study (Fariss 2013). Third, I discuss the area of disagreement between the new latent variable estimates and the estimates from Fariss (2013) and why this disagreement occurs. Fourth, I discuss general trends in the new count data over time, which corroborate recent findings of a decline in the number of fatalities during war time (Goldstein 2011; Lacina, Gleditsch and Russett 2006) and a decline in the level of violence more generally (Pinker 2011). Finally, I demonstrate how the new count estimates strengthen the inference from an existing study of UN peace keeping interventions (Hultman 2013) by accounting for the censoring of low levels of government killings.

4.5.1 Model Comparisons

Here I present correlation coefficients between the three observed one-sided government killing count variables $\{Best, Low, High\}$ and the estimated count variable from two latent variable models. Model M_5 assumes one over-dispersion parameter r_j for each of the three one-sided government killing variables. Model M_6 assumes a unique over-dispersion parameter r_{it} for each of the observations. Model M_6 is the preferred model both because of the evidence from the correlation coefficients but also because it relaxes the assumption that the level of one-sided government killing can be uniformly observed across units. If this assumption held, then M_5 would be a reasonable model. However, as the UCDP coders acknowledge, this is an untenable assumption when attempting to quantify the number of repressive events, which governments have an incentive to hide.

Table 4.3: Correlation coefficients between the three observed one-sided government killing count variables $\{Best, Low, High\}$ and the estimated count variable from two latent variable models. Model M_5 assumes one over-dispersion parameter r_j for each of the three one-sided government killing variables. Model M_6 assumes a unique over-dispersion parameter r_{it} for each of the observations.

	ρ_{M_5}	ρ_{M_6}
Best	0.693 [0.620, 0.755]	0.919 [0.897, 0.937]
Low	0.746 [0.683, 0.798]	0.933 [0.914, 0.948]
High	0.710 [0.633, 0.763]	0.909 [0.883, 0.929]

Figure 4.1, Figure 4.2, and Figure 4.3 present the point estimates from the observed UCDP count variable plotted against the predicted posterior count from the latent variable model (M_6) for the Best, Low, and High counts respectively. The correlation between the observed count variable and the estimated count variable are all relatively high as displayed in the second column of the Table above. The weight of the distribution falls below the 45 degree line, which suggests that the best estimate from the UCDP conflict dataset is a conservative one. This result accords with the "general rule" for UCDP's estimation of one-sided violence, which is "moderation" (see the appendix, which quotes the UCDP codebook).

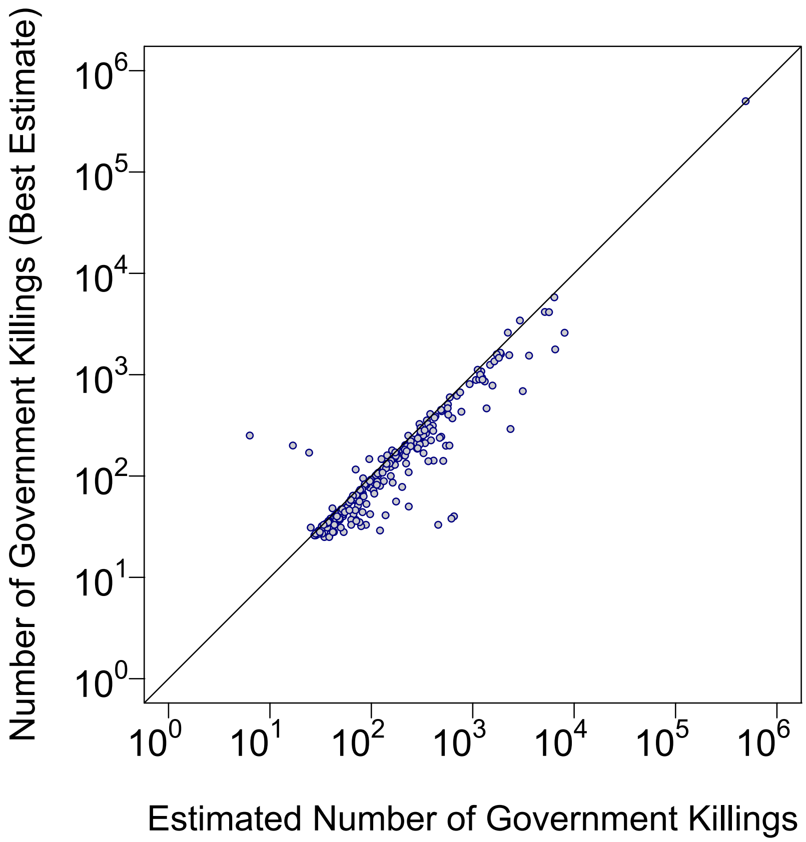


Figure 4.1: The points are the observed best value for the count variable plotted against the predicted posterior count from the latent variable model. The correlation between the observed count variable and the estimated count variable is $\rho = 0.919$ $[0.897, 0.937]$.

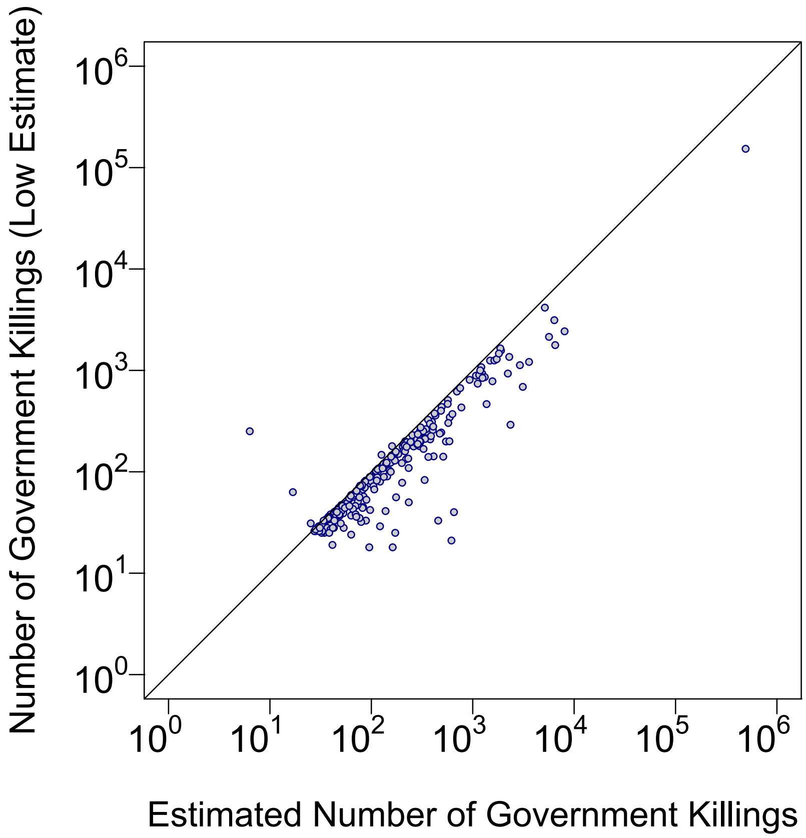


Figure 4.2: The points are the observed low value for the count variable plotted against the predicted posterior count from the latent variable model. The correlation between the observed count variable and the estimated count variable is $\rho = 0.909[0.883, 0.929]$.

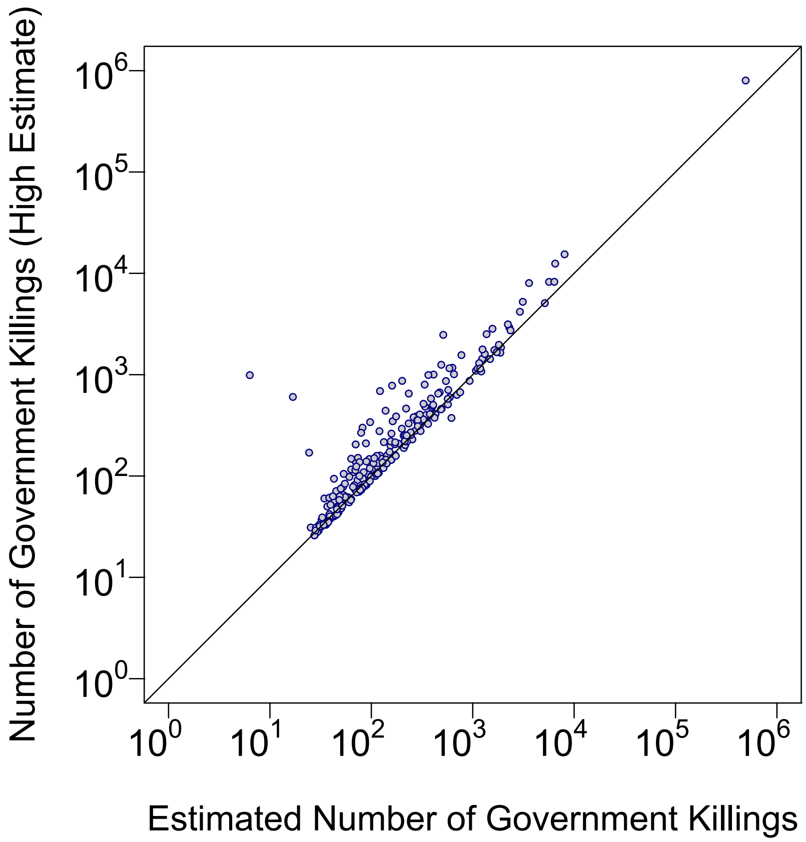


Figure 4.3: The points are the observed high value for the count variable plotted against the predicted posterior count from the latent variable model. The correlation between the observed count variable and the estimated count variable is $\rho = 0.933[0.914, 0.948]$.

4.5.2 New Latent Repression Variable Estimates (With Counts) vs. Existing Latent Variable Estimates (Without Counts)

Figure 4.4 captures the increasing disagreement between the latent variables estimates generated from the dynamic standard model and those from the constant standard model (1976-2010) estimated in the models presented by Fariss (2013). The disagreement occurs because the dynamic standard model incorporates the changing standard of accountability, whereas the constant standard model does not. Bias results when temporal changes to the documentary sources used to generate standards based data are not accounted for. The disagreement is clear visually even though the correlation between the point estimates from the two models is quite high ($\rho = 0.961$). Figure 4.5 captures the same pattern. This result corroborates the new model presented in this paper compared with the original constant standard model with virtually the same correlation coefficient ($\rho = 0.961$).

Figure 4.6 shows both the high level of agreement between the original dynamic standard model (no event-based count data) and the new version of this model that incorporates event-based count data. The correlation is even higher ($\rho = 0.995$) than those from the other figures. Disagreement between the two latent variable estimates occurs in the region from approximately the mean level of the original dynamic standard variable to approximately -2.0 standard deviations below the mean value. This is the region along the latent variable at which the UCDP event-count data begin to be recorded. I explore the implications for these disagreements in a replication analysis, which I present below. Before presenting this replication however, I explore the new count estimates in the region of disagreement highlighted in Figure 4.6.

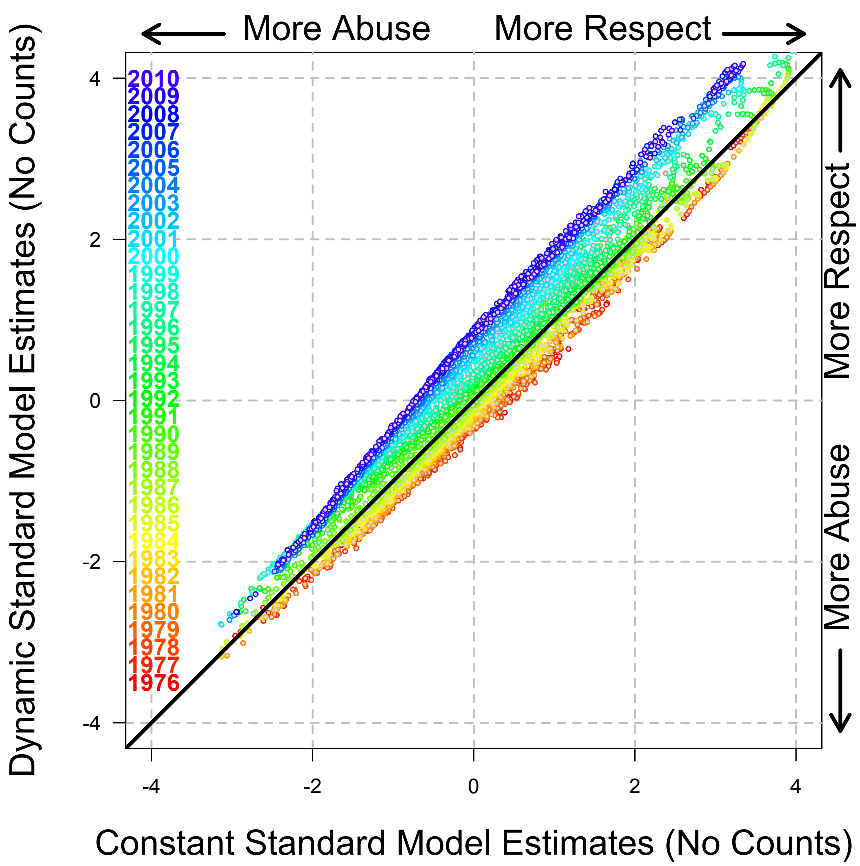


Figure 4.4: Relationship between the latent variable estimates generated from the Dynamic Standard Model (no event-based count data) on the y-axis and the estimates generated from the Constant Standard Model (no event-based event data) on the x-axis (1976-2010). The 45-degree line represents perfect agreement between the two estimates. Disagreement between the two sets of estimates increases as a function of time.

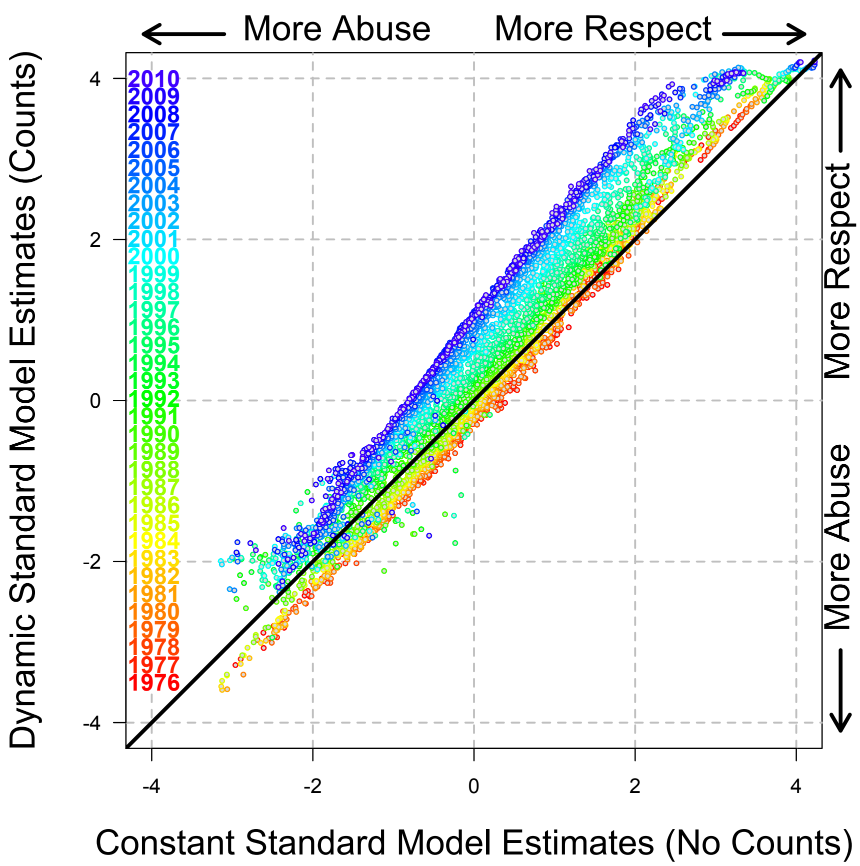


Figure 4.5: Relationship between the latent variable estimates generated from the Dynamic Standard Model (event-based count data) on the y-axis and the estimates generated from the Constant Standard Model (no event-based event data) on the x-axis (1976-2010). The 45-degree line represents perfect agreement between the two estimates. Disagreement between the two sets of estimates increases as a function of time and is consistent with the relationship displayed in Figure 4.4.

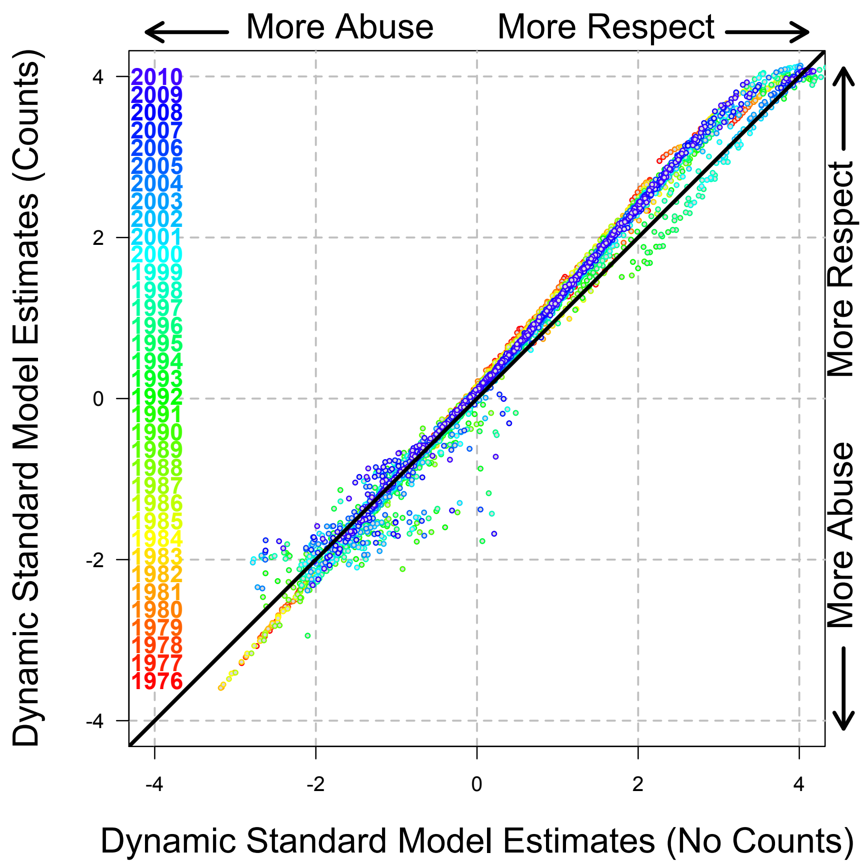


Figure 4.6: Relationship between the latent variable estimates generated from the Dynamic Standard Model (event-based count data) on the y-axis and the estimates generated from the Dynamic Standard Model (no event-based event data) on the x-axis (1976-2010). Disagreement between the two latent variable estimates occurs in the region from approximately the mean level of the original dynamic standard variable to approximately -2.0 standard deviations below the mean value. This is the region along latent variable at which the UCDP event-count data begin to be recorded.

4.5.3 Model Predictions Using the New Latent Repression Variable

New latent variable estimates of repression provide predictions of the expected number of government one-sided killings. Figure 4.7, Figure 4.8, and Figure 4.9 display the model predictions and observed data (Best, Low, and High counts respectively). The red line in each of these figures represents the posterior expectation of the count variable as a function of the latent repression variable. The points are the observed value (Best, Low, and High counts respectively in the three figures) reported by the UCDP conflict dataset for the count variable plotted against the corresponding latent variable estimate. The model suggests that one-sided government killing stops at approximately the mean value of the "true" level of repression. Government-killing reaches a level great enough for the UCDP conflict coders to find sufficient evidence of such human rights violations such that the country-year observations enter the dataset at approximately one standard deviation (-1.0) below the mean value of the "true" level of repression. The magnitude of the predictions increases as the latent variable decreases. Though, only Rwanda (1994) nears the maximum observed value, the model makes predictions that accord with earlier episodes of domestic political violence that occurred prior to 1989 when the coverage of the UCDP conflict dataset begins. I discuss these patterns in more detail in the next section.

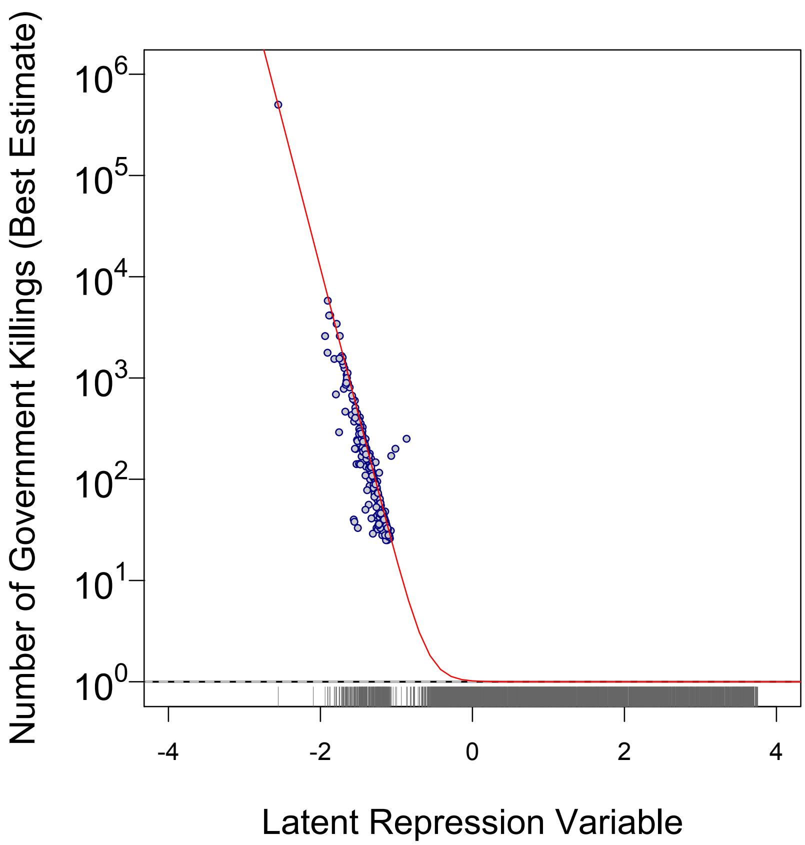


Figure 4.7: New latent variable estimates of repression provide predictions of the expected number of government one-sided killings (Best). The red line is the posterior expectation of the count variable. The points are the observed best value reported by the UCDP conflict dataset for the count variable plotted against the corresponding latent variable estimate.

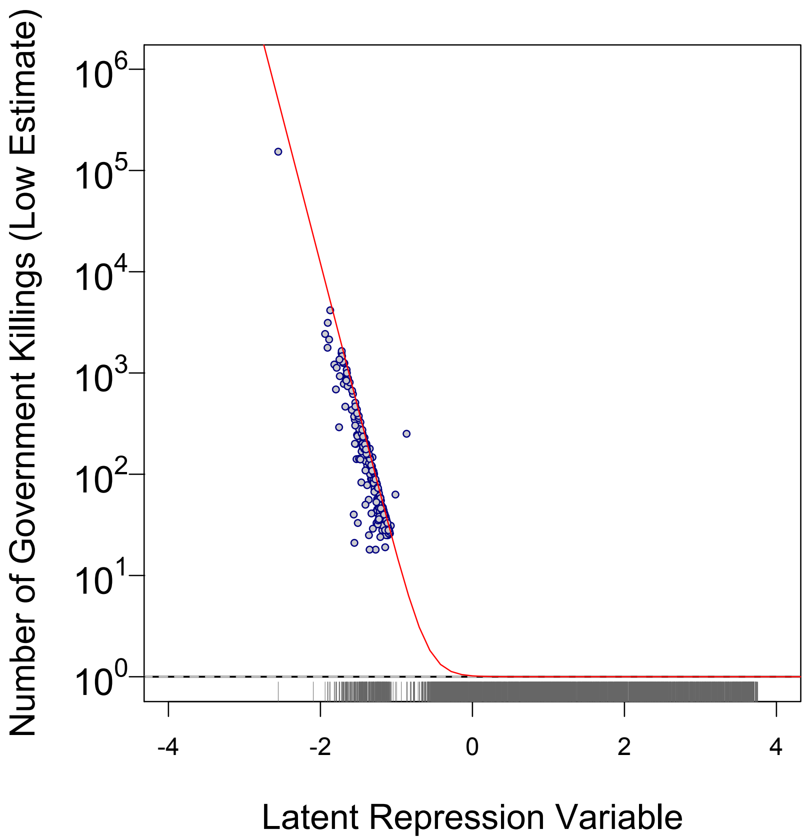


Figure 4.8: New latent variable estimates of repression provide predictions of the expected number of government one-sided killings (Low). The red line is the posterior expectation of the count variable. The points are the observed best value reported by the UCDP conflict dataset for the count variable plotted against the corresponding latent variable estimate.

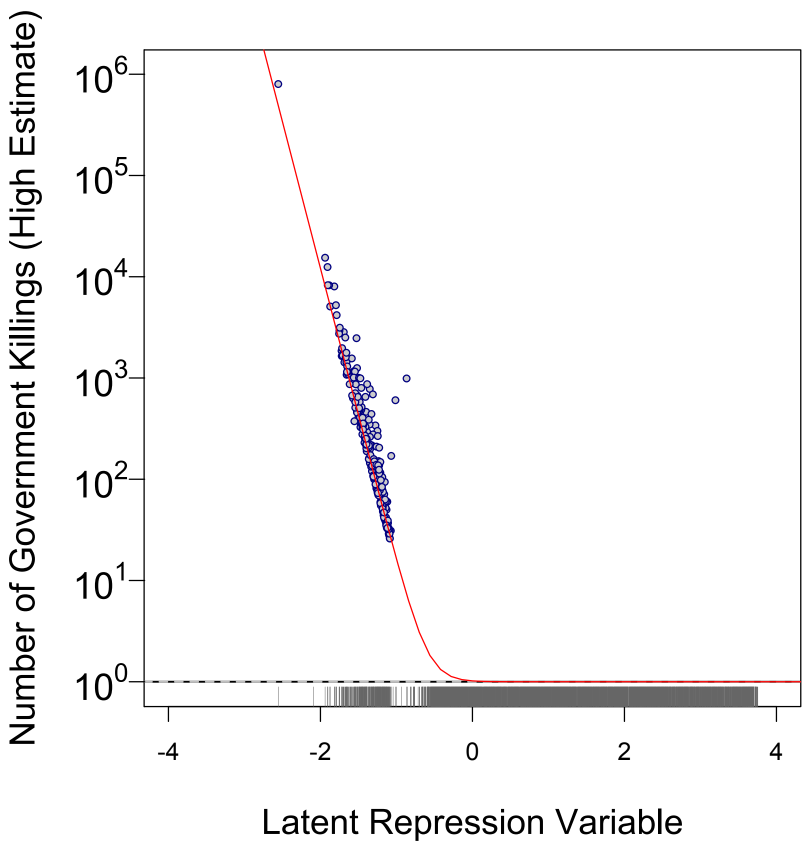


Figure 4.9: New latent variable estimates of repression provide predictions of the expected number of government one-sided killings (High). The red line is the posterior expectation of the count variable. The points are the observed best value reported by the UCDP conflict dataset for the count variable plotted against the corresponding latent variable estimate.

4.5.4 Changes in Government-Killing Over Time

To get a better sense of the overall pattern of government killing over time consider the following figures. The left panels in Figure 4.10 display the total of the new count estimates for all observations (1989-2010) compared to total estimates from the original UCDP data (Best and High). The right panels in Figure 4.10 display the total of the new count estimates for only those observations contained in the original UCDP data again compared to the total of the original UCDP data. These comparisons provide information about the level of agreement between the original data and the new estimates. Note that the estimated total and UCDP total are in close agreement in the right hand panels ($\rho = 0.916$ $[0.807, 0.965]$ in the upper right panel and $\rho = 0.924$ $[0.824, 0.968]$ in the lower right panel). The increased disagreement occurs in the left panels ($\rho = 0.692$ $[0.382, 0.862]$ in the upper left panel and $\rho = 0.695$ $[0.386, 0.863]$ in the lower left panel) because of the additional observations that are now no longer assumed to be 0 and added to the estimated total. Disagreement — the upward shift away from the 45-degree line in the left panels — is greatest in the earliest years of the UCDP data (i.e., 1989, 1990, and 1991). The results for these early years suggest that information about government one sided killing was more easy to obtain after the end of the Cold War and the breakup of the former Soviet Union than during this period.

The latent variable model provides estimated totals that go back to the beginning of the series in 1949. Figure 4.11 displays the total number of one-sided killings each year for the entire period. Readers should keep in mind that these estimate are not based on count data from 1949 until 1988. However, event-based information contained in several of the binary indicators is included in the model (see the appendix for details). These estimated totals are an approximation of the overall level of state sanctioned government killing directed at civilians. The model suggests that prior to the end of the Cold War, more than a million one-sided government killings occurred each year. The number dropped into the high 1000s during the 1990s (other than during the Rwandan genocide in 1994) and most recently to just below 1000. Though these estimates are likely conservative because of the coding procedures adopted by UCDP project, the estimates nonetheless corroborate the results from other studies that find a similar decline in the number of fatalities during war time (Goldstein 2011; Lacina, Gleditsch and Russett 2006) and a decline in the level of violence more generally (Pinker 2011). All of these authors point out that the decline in violence has in no way been a steady one. And sadly, the estimated totals for the next few years of data will rise because of the ongoing civil conflict in Syria.

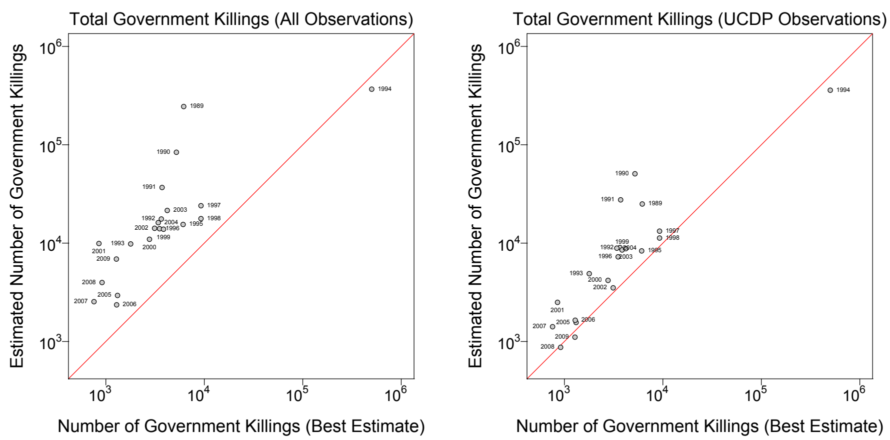


Figure 4.10: Model based estimates of the yearly total of one-sided government killings compared to the reported total from the UCDP data (Best and High estimates). Note that the estimated total and UCDP total are in close agreement in the right hand panels ($\rho = 0.916$ $[0.807,\ 0.965]$ in the upper right panel and $\rho = 0.924$ $[0.824,\ 0.968]$ in the lower right panel). The increased disagreement occurs in the left panels ($\rho = 0.692$ $[0.382,\ 0.862]$ in the upper left panel and $\rho = 0.695$ $[0.386,\ 0.863]$ in the lower left panel) because of the additional observations that are now no longer assumes to be 0 and added to the estimated total.

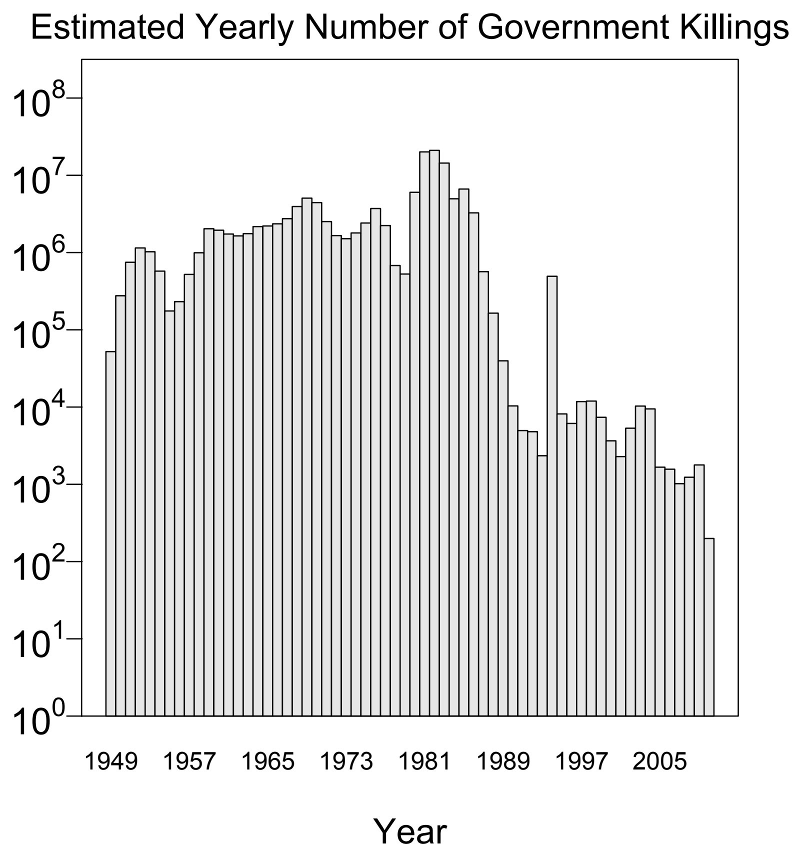


Figure 4.11: Model based estimates of the yearly number of one-sided government killings beginning in 1949 and ending in 2010. These estimated totals corroborate the results from other studies that find a similar decline in the number of fatalities during war time (Goldstein 2011; Lacina, Gleditsch and Russett 2006) and a decline in the level of violence more generally (Pinker 2011).

4.5.5 Replication

Here I replicate a recent analysis to demonstrate the the substantive importance of the new estimates of one-sided government killing. In a recent issue of the *Journal of Peace Research*, Hultman (2013) demonstrated an empirical link between United Nations interventions and the level of one-sided killing committed by government agents during civil wars. The replication analysis presented here demonstrates that the relationship is even stronger than the one found in the original article. This result obtains because of the additional cases of low level one-sided killing that the new estimates provide. In the original analysis, civil-war-years that did not have an estimate of one-sided government killing greater than 25 were assumed to be 0. The new estimates for these units strengthens the original results as the positive differences between the logistic regression coefficients displayed in Table 4.4 indicate.

Table 4.4: Replication of study using one-sided government killing data.

	Original	Replicated	Difference	$p-value$	n
Model 1	0.164 (0.060)	0.450 (0.151)	0.286	0.039	850
Model 2	0.379 (0.158)	0.418 (0.231)	0.039	0.445	890
Model 3	0.400 (0.180)	1.013 (0.346)	0.613	0.058	850
Model 4	0.684 (0.260)	0.611 (0.381)	-0.073	0.437	890
Model 5	0.569 (0.190)	1.074 (0.347)	0.505	0.101	345

The difference between the coefficients from the replicated models and from the original models (main effect) is $\beta_{replication} - \beta_{original}$ and displayed in Table 4.4. The p-value for this difference is simply based on the following $Z-score$: $\frac{\beta_{replication}-\beta_{original}}{\sqrt{SE(\beta_{replication})^2+SE(\beta_{original})^2}}$. Though only some of these differences are statistically significant at conventional levels, the substantive importance is clear. The uncensored data generated from the extended latent variable model have strengthened the original relationship between violence against civilians and UN interventions found by Hultman (2013). Analysts that use the UCDP one-sided government killing data for future research should consider using the new count estimates presented in this paper along side the three existing estimates. Inconsistent results would suggest that censoring is biasing the results between models using the original data and estimates using the new data presented here. In the case of the findings presented by Hultman (2013), the censoring is biasing the coefficients closer towards 0 for three of the five models.

4.6 Conclusion

The model developed in this paper allows for event-based count data to help improve the estimation of the latent repression estimates developed by Schnakenberg and Fariss (Forthcoming) and extended by Fariss (2013). The model goes beyond this however, by providing new count estimates of the expected number of one sided government killings. The new estimates enhance the existing event-based government-killing data by providing estimates for country-years that are otherwise assumed to be

0 in applied research. Moreover, the new estimates strengthen existing findings that relate the level of one-sided government killing with UN interventions (Hultman 2013).

The new count estimates quantify the uncertainty inherent in the estimates of event-based count data by providing information about the underlying distribution from which such observations are drawn. I demonstrated how the new latent variable estimates confirm the inferences generated by Fariss (2013) and that the new count data predicts the existing event-based count estimates. Thus, the new count estimates represent a more accurate and informative "representation of reality" (Sundberg 2009, pg. 4). As Sundberg states in one of the many UCDP data reports, "[e]ven though the best open-source information available is commonly used by the coders the figures given are still only estimates. They should be viewed as a representation of reality only, and not an attempt to reflect the true number of fatalities" (Sundberg 2009, pg. 4). The models developed here parameterize the inherent uncertainty in quantifying the repressive actions of governments.

The extensions of the latent variable model presented here and by others (Schnakenberg and Fariss Forthcoming; Fariss 2013) represent only some of the first steps toward improving the measurement and understanding of repressive behaviors. Overall, researchers and activists want to make inferences about more than just country-year units of analysis. New data collection efforts are beginning to acknowledge and understand the role of different state actors who commit human rights violations and the different groups that are targeted. To date, the ITT data project (Conrad and Moore 2011; Conrad, Haglund and Moore 2012), and UCDP data project (Eck and Hultman 2007; Sundberg 2009) are the only data efforts that systematically collect repression data about targets, agents, or non-state actors for all states. Other event-based data collection efforts exist and are also beginning to provide some of this information for specific regions (i.e., Salehyan and Hendrix 2012; Saleyhan et al. 2012).[6]

The models presented in this paper are capable of systematically linking these new and diverse sources of information with existing categorical data. The modeling techniques available allow for the incorporation of information at multiple levels in one model (e.g., country-year, country-year-actors, country-year-victims, country-year-regions). Moreover, these models can begin to link together a great variety of data from historical sources such as the level of political imprisonment in Soviet gulags (Getty, Rittersporn and Zemskov 1993), the number of death squad killings during the civil war in El Salvador (Mason and Krane 1989; Mason 1999), or the number of disappeared in Latin America as documented by several truth commission reports (Clark and Sikkink Forthcoming; Sikkink 2011). The hierarchical models and computational resources necessary to construct and estimate these models are available and ready to be exploited to improve our understanding of repression. Importantly, the models I have developed here allow for the acknowledgment and quantification of disagreements between different sources of information, which should assuage concerns from some researchers still skeptical of models that compare event counts from disparate sources of information.[7]

[6] For more a thorough review of event-based data in international relations and comparative politics see the recent special issue of *International Interactions* "New Event Data in Conflict Research", Volume 38, Issue 4. See especially the final entry in the volume by Schrodt (2012).

[7] See Poe (2004) for a short review of the debate about the appropriateness of comparing event-based data.

4.7 Appendix

4.7.1 UCDP One-sided Violence Dataset

Best, Low and High fatality estimates from Eck and Hultman (2007) data. The following passage is quoted directly from the most recent code book (UCDP One-sided Violence Dataset v 1.4-2012, 1989-2011).

> The general rule for UCDP's estimation of one-sided violence is moderation. All incidents have to be verified in one way or another, and all estimates reported are based on UCDP expertise of each particular conflict. As a general rule, all figures are disaggregated as far as possible and any figures that are not trustworthy are disregarded in the coding process. Due to the great uncertainty of reports from conflict areas, the project provides three estimates concerning battle - related deaths for each year.
>
> (a) Best estimate. The UCDP Best estimate consist of the aggregated most reliable numbers for all incidents of one - sided violence during a year. If different reports provide different estimates, an examination is made as to what source is most reliable. If no such distinction can be made, UCDP as a rule include the lower figure given.
>
> (b) Low estimate. The UCDP Low estimate consists of the aggregated low estimates for all incidents of one - sided violence during a year. If different reports provide different estimates and a higher estimate is considered more reliable, the low estimate is also reported if deemed reasonable.
>
> (c) High estimate. The UCDP High estimate consists of the aggregated high estimates for all incidents of one - sided violence during a year. If different reports provide different estimates and a lower estimate is considered more or equally reliable, the high estimate is also reported if deemed reasonable. If there are incidents when there is some uncertainty about what party have been involved , these may also be included in the high estimate.

4.7.2 Other Data Sources

Table 4.5 and Table 4.6 contain information about the documentary sources used to generate each of the variables that enter the latent variable models presented in this paper. For more information on these sources see the original citations and also Fariss (2013).

Table 4.5: Standards-Based Repression Data Sources

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
CIRI Physical Integrity Data, 1981-2010 - political imprisonment (ordered scale, 0-2) - torture (ordered scale, 0-2) - extrajudicial killing (ordered scale, 0-2) - disappearance (ordered scale, 0-2)	Cingranelli and Richards (1999, 2012*a*,*b*) Amnesty International Reports[1] and State Department Reports[2] *Information in Amnesty reports takes precedence over information in State Department reports*
Hathaway Torture Data, 1985-1999 - torture (ordered scale, 1-5)	Hathaway (2002) State Department Reports[1]
Ill-Treatment and Torture (ITT), 1995-2005 - torture (ordered scale, 0-5)	Conrad and Moore (2011), Conrad, Haglund and Moore (2012), Amnesty International (2006) Annual Reports[1], press releases[1], and Urgent Action Alerts[1]
PTS Political Terror Scale, 1976-2010 - Ammesty International scale (ordered scale, 1-5) - State Department scale (ordered scale, 1-5)	Gibney, Cornett and Wood (2012), Gibney and Dalton (1996) Amnesty International Reports[1] State Department Reports[1]

1. Primary Source; 2. Secondary Source

Table 4.6: Event-Based Repression Data Sources

Dataset Name and Variable Description	**Dataset Citation** and Primary Source Information
Harff and Gurr Dataset, 1946-1988 - massive repressive events (1 if country-year experienced event 0 otherwise)	Harff and Gurr (1988) historical sources (see article references)[1]
Political Instability Task Force (PITF), 1956-2010 - genocide and politicide (1 if country-year experienced event 0 otherwise)	Harff (2003), Marshall, Gurr and Harff (2009) historical sources (see article references)[1] State Department Reports[2] Amnesty International Reports[2]
Rummel Dataset, 1949-1987 - genocide and democide (1 if country-year experienced event 0 otherwise) (3 death count estimates: best, low, high)	Rummel (1994*b*, 1995), Wayman and Tago (2010) New York Times[1], New International Yearbook[2], Facts on File[2], Britannica Book of the Year[2], Deadline Data on World Affairs[2], Kessing's Contemporary Archives[2]
UCDP One-sided Violence Dataset, 1989-2010 - government killing (event count estimate) (1 if country-year experienced event 0 otherwise) (3 death count estimates: best, low, high)	Eck and Hultman (2007), Sundberg (2009) Reuters News[1], BBC World Monitoring[1] Agence France Presse[1], Xinhua News Agency[1], Dow Jones International News[1], UN Reports[2], Amnesty International Reports[2], Human Rights Watch Reports[2], local level NGO reports (not listed)[2]
World Handbook of Political and Social Indicators WHPSI, 1948-1982 - political executions (event count estimate) (1 if country-year experienced event 0 otherwise)	Taylor and Jodice (1983) New York Times[1], Middle East Journal[2], Asian Recorder[2], Archiv der Genenwart[2] African Diary[2], Current Digest of Soviet Press[2]

1. Primary Source; 2. Secondary Source

4.8 Acknowledgements

I would like to thank Jesse Driscoll, James Fowler, Miles Kahler, David Lake, Keith Schnakenberg, and Brice Semmens for many helpful comments and suggestions. The estimates from this paper along with the code necessary to implement the models in JAGS and R will be made publicly available at: `http://dvn.iq.harvard.edu/dvn/dv/HumanRightsScores`. This chapter is in preparation for submission to a peer-reviewed political science journal.

Bibliography

Aleman, Eduardo and Sebastián M. Saiegh. 2007. "Legislative Preferences, Political Parties, and Coalition Unity in Chile." *Comparative Politics* 39(3):253–272.

Amnesty International. 2006. *Amnesty International's Country Dossiers and Publications, 1962-2005*. Leiden: IDC Publishers.
URL: *http://www.idcpublishers.com/ead/ead.php?faid=127faid.xml*

Armstrong, David A. 2011. "Stability and change in the Freedom House political rights and civil liberties measures." *Journal of Peace Research* 48(5):653–662.

Banks, Arthur S. and Thomas C. Muller. 1998. *Political Handbook of the World*. Binghamton, New York: CSA Publications.

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.

Bell, Sam, K. Chad Clay and Amanda Murdie. 2012. "Neighborhood Watch: Spatial Effects of Human Rights INGOs." *Journal of Politics* 74(2):354–368.

Bennett, D. Scott and Timothy Nordstrom. 2000. "Foreign Policy Substitutability and Internal Economic Problems in Enduring Rivalries." *Journal of Conflict Resolution* 44(1):33–61.

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

Bolker, Benjamin M. 2008. *Ecological Models and Data in R*. Princeton, NJ: Princeton University Press.

Bollen, Kenneth A. 1986. "Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984." *Human Rights Quarterly* 8(4):567–591.

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489(7415):295–298.

Borsboom, Denny. 2005. *Measuring the Mind*. Cambridge: Cambridge University Press.

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

Bueno De Mesquita, Bruce, Feryal M. Cherif, George W. Downs and Alastair Smith. 2005. "Thinking Inside the Box: A Closer Look at Democracy and Human Rights." *International Studies Quarterly* 49(3):439–458.

Butts, Carter T. 2012. "The SNA Package, v2.2.".

Cameron, A. Colin and Pravin K. Trivedi. 1986. "Econometric models based on count data: comparisons and applications of some estimators and tests." *Journal of Applied Econometrics* 1:29–54.

Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Carey, Sabine C. 2006. "The Dynamic Relationship Between Protest and Repression." *Political Research Quarterly* 59(1):1–11.

Carey, Sabine C. 2007. "Rebellion in Africa: Disaggregating the effect of political regimes." *Journal of Peace Research* 44(1):47–64.

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.

Central Intelligence Agency (CIA). 2004. "CIA world factbook.".
URL: *http://www.cia.gov/cia/publications/factbook*

Christakis, Nicholas A. and James H. Fowler. 2008. "The Spread of Obesity in a Large Social Network Over 32 Years." *The New England Journal of Medicine* 357(4):370–379.

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*

Cingranelli, David L. and Mikhail Filippov. 2010. "Electoral Rules and Incentives to Protect Human Rights." *Journal of Politics* 72(1):243–257.

Cioffi-Revilla, Claudio and Harvey Starr. 1995. "Opportunity, Willingness and Political Uncertainty: Theoretical Foundations of Politics." *Journal of Theoretical Politics* 7(4):447–476.

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

Clark, Ann Marie and Kathryn Sikkink. Forthcoming. "Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures?" *Human Rights Quarterly* TBD.

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

Conrad, Courtenay R. and Jacqueline H.R. Demeritt. 2011. "Options in the Arsenal: Are Repressive Tactics Substitutes or Complements?" *Presented at the Annual Meeting of the American Political Science Association* .

Conrad, Courtenay R., Jillienne Haglund and Will H. Moore. 2012. "Disaggregating Torture Allegations: Introducing the Ill-Treatment and Torture (ITT) Country-Year Data." *International Studies Perspectives* forthcoming.

Conrad, Courtenay R. and Will H. Moore. 2010. "What Stops the Torture?" *American Journal of Political Science* 54(2):459–476.

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*

Cranmer, Skyler J., Bruce A. Desmarais and Justin H. Kirkland. 2012. "Toward a Network Theory of Alliance Formation." *International Interaction* 38(3):295–324.

Cranmer, Skyler J., Tobias Heinrich and Bruce A. Desmarais. Forthcoming. "Reciprocity and the Structural Determinants of the International Sanctions Network." *Social Networks* .

Dancy, Geoff and Kathryn Sikkink. 2012. "Ratification and Human Rights Prosecutions: Toward a Transnational Theory of Treaty Compliance." *NYU Journal of International Law and Politics* 44(3):751–790.

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.

Davenport, Christian. 2007*a*. "State repression and political order." *Annual Review of Political Science* 10:1–23.

Davenport, Christian. 2007*b*. "State repression and the tyrannical peace." *Journal of Peace Research* 44(4):485–504.

Davenport, Christian. 2009. *Media Bias, Perspective, and State Repression: The Black Panther Party*. Cambridge, MA: Cambridge University Press.

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

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

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.

Demeritt, Jacqueline H.R. 2012. "International Organizations and Government Killing: Does Naming and Shaming Save Lives?" *International Interactions* 38(5):597–621.

Desposato, Scott W. 2006. "The impact of electoral rules on legislative parties: Lessons from the Brazilian Senate and Chamber of Deputies." *Journal of Politics* 68(4):1018–1030.

Donnelly, Jack. 2003. *Universal Human Rights*. Ithaca, NY: Cornell University Press.

Donnelly, Jack and Rhoda E. Howard. 1988. "Assessing National Human Rights Performance: A Theoretical Framework." *Human Rights Quarterly* 10(2):214–248.

Downs, George W., David M. Rocke and Peter N. Barsoom. 1996. "Is the good news about compliance good news about cooperation?" *International Organization* 50(3):379–406.

Duvall, Raymond D. and Michael Stohl. 1983. Governance by Terror. In *The Politics of Terrorism*, ed. Michael Stohl. New York: Marcel Dekker pp. 179–219.

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.

Fariss, Christopher J. 2010. "The Strategic Substitution of United States Foreign Aid." *Foreign Policy Analysis* 6(2):107–131.

Fariss, Christopher J. 2013. "Respect for Human Rights has Improved Over Time: A Dynamic Latent Variable Model." *UCSD Working Paper* .

Fariss, Christopher J. and Keith Schnakenberg. 2013. "Measuring Mutual Dependence Between State Repressive Actions." *Journal of Conflict Resolution* forthcoming.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):75–90.

Fiske, Susan T., Lasana T. Harris and Amy J. C. Cuddy. 2004. "Why Ordinary People Torture Enemy Prisoners." *Science* 306(5701):1482 – 1483.

Fowler, James H. 2006. "Connecting the Congress: A Study of Cosponsorship Networks." *Political Analysis* 14:456–487.

Fowler, James H. and Cindy D. Kam. 2008. "Beyond the Self: Social Identity, Altruism, and Political Participation." *Journal of Politics* 69(3):813–827.

Fowler, James H., Timothy R. Johnson, James F. Spriggs II, Sangick Jeon and Paul J. Wahlbeck. 2007. "Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court." *Political Analysis* 15:324–346.

Freedom House. 2009. "Freedom in the World Historical Rankings.".
URL: *http://www.freedomhouse.org*

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.

Getty, J. Arch, Gábor T. Rittersporn and Viktor N. Zemskov. 1993. "Victims of the Soviet Penal System in the Pre-War Years: A First Approach on the Basis of Archival Evidence." *The American Historical Review* 98(4):1017–1049.

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.

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

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed conflict 1946–2001: A new dataset." *Journal of Peace Research* 39(5):615–637.

Goldstein, Joshua S. 2011. *Winning the War on War: The Decline of Armed Conflict Worldwide*. New York, NY: Dutton.

Goldstein, Robert Justin. 1978. *Political Repression in Modern America, From 1870 to Present*. Cambridge, MA: G. K. Hall.

Goodman, Ryan and Derek Jinks. 2003. "Measuring the Effects of Human Rights Treaties." *European Journal of International Law* 14(1):171–183.

Greene, William. 2008. "Functional forms for the negative binomial model for count data." *Economic Letters* 99:585–590.

Guttman, Louis. 1949. *The basis for scalogram analysis*. Indianapolis, Ind. : Bobbs-Merrill.

Hafner-Burton, Emilie M. and James Ron. 2009. "SEEING DOUBLE Human Rights Impact through Qualitative and Quantitative Eyes." *World Politics* 61(2):360–401.

Hafner-Burton, Emilie M. and Kiyoteru Tsutsui. 2005. "Human rights in a globalizing world: The paradox of empty promises." *American Journal of Sociology* 110(5):1373–1411.

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

Harbom, Lotta, Erik Melander and Peter Wallensteen. 2008. "Dyadic Dimensions of Armed Conflict, 1946-2007." *Journal of Peace Research* 45(5):697–710.

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.

Henderson, Conway W. 1991. "Conditions Affecting the Use of Political Repression." *Journal of Conflict Resolution* 35(1):120–142.

Hidalgo, C. A., B. Klinger, A. L. Barabási and R. Hausmann. 2007. "The Product Space Conditions the Development of Nations." *Science* 317:482–487.

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

Hill Jr., Daniel W., Will H. Moore and Bumba Mukherjee. Forthcoming. ""Information Politics v Organizational Incentives: When are Amnesty International's "Naming and Shaming" Reports Biased?" *International Studies Quarterly* .

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

Hopgood, Stephen. 2006. *Keepers of the Flame: Understanding Amnesty International*. Ithaca: Cornell University Press.

Hultman, Lisa. 2013. "UN peace operations and protection of civilians: Cheap talk or norm implementation?" *Journal of Peace Research* 50(1):59–73.

Imai, Kosuke, Gary King and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software.".
URL: *http://GKing.Harvard.Edu/zelig*

Innes de Neufville, Judith. 1986. "Human Rights Reporting as a Policy Tool: An Examination of the State Department Country Reports." *Human Rights Quarterly* 8(4):681–699.

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

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

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. New York, NY: John Wiley and Sons.

Jones, Jason J. 2012. "Scale-Free Superiority, Egocentric Bias and Network Centrality Heuristics in Social Graph Learning." *working paper* .

Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow and James H. Fowler. 2012. "Inferring Tie Strength from Online Directed Behavior." *PLoS One* 8(1):e52168.

Kahler, Miles, ed. 2009. *Networked Politics: Agency, Power, and Governance*. Ithaca, NY: Cornell University Press.

Kamada, T. and S. Kawai. 1989. "An algorithm for drawing general undirected graphs." *Information Processing Letters* 31(1):7–15.

Keck, Margaret and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.

Keith, Linda Camp. 1999. "The United Nations international covenant on civil and political rights: Does it make a difference in human rights behavior?" *Journal of Peace Research* 36(1):95–118.

Keith, Linda Camp, C. Neal Tate and Steven C. Poe. 2009. "Is The Law A Mere Parchment Barrier To Human Rights Abuse?" *Journal of Politics* 71(1):644–660.

Kim, Hunjoon and Kathryn Sikkink. 2010. "Explaining the Deterrence Effect of Human Rights Prosecutions for Transitional Countries." *International Studies Quarterly* 54(4):939–963.

King, Gary. 1989. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly* 33(2):123–147.

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.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):341–355.

Korey, William. 2001. *NGOs and the Universal Declaration of Human Rights: a curious grapevine*. Palgrave Macmillan.

Krain, Matthew. 2012. "J'accuse! Does Naming and Shaming Perpetrators Reduce the Severity of Genocides or Politicides?" *International Studies Quarterly* .

Lacina, Bethany, Nils Petter Gleditsch and Bruce M. Russett. 2006. "The Declining Risk of Death in Battle." *International Studies Quarterly* 50(3):673–680.

Lake, David A. and Wendy Wong. 2009. The Politics of Networks: Interests, Power, and Human Rights Norms. In *Networked Politics*, ed. Miles Kahler. Ithaca, NY: Cornell University Press.

Landman, Todd. 2004. "Measuring Human Rights: Principle, Practice, and Policy." *Human Rights Quarterly* 26(4):906–931.

Landman, Todd. 2005. "The Political Science of Human Rights." *British Journal of Political Science* 35(3):549–572.

Landman, Todd and Marco Larizza. 2009. "Inequality and Human Rights: Who Controls What, When, and How." *International Studies Quarterly* 53(3):715–736.

Lawyers Committee for Human Rights. 1997. *Critique: Review of the U.S. Department of State's Country Reports on Human Rights Practices for 1996*. University of Pennsylvania Press.

Leiter, Karen, Voravit Suwanvanichkij, Ingrid Tamm, Vincent Iacopino and Chris Beyrer. 2006. "Human rights abuses and vulnerability to HIV/AIDS: the experiences of Burmese women in Thailand." *Health and Human Rights* 9(2):88–111.

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.

Lord, Frederic M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum Associates.

Lord, Frederic M. and Melvin R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Addison-Wesley.

Lupu, Yonatan. 2013*a*. "Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements." *International Organization* Forthcoming.

Lupu, Yonatan. 2013*b*. "The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects." *American Journal of Political Science* Forthcoming.

Lupu, Yonatan and Erik Voeten. 2012. "Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights." *British Journal of Political Science* 42(2):413–439.

Lupu, Yonatan and James H. Fowler. 2012. "Strategic Citations to Precedent on the U.S. Supreme Court." *working paper* .

Lupu, Yonatan and Vincent Traag. 2013. "Trading Communities, the Networked Structure of International Relations and the Kantian Peace." *Journal of Conflict Resolution* .

Madani, Hamed. 1992. Socioeconomic development and military policy consequences of Third World military and civilian regimes, 1965-1985 PhD thesis University of North Texas Denton, TX: .

Maoz, Zeev. 2009. "The Effects of Strategic and Economic Interdependence on International Conflict Across Levels of Analysis." *American Journal of Political Science* 53(1):223–240.

Marshall, Monty G., Ted R. Gurr and Barbara Harff. 2009. "PITF - STATE FAILURE PROBLEM SET: Internal Wars and Failures of Governance, 1955-2009." *Dataset and Coding Guidelines* .

Marshall, Monty, Keith Jaggers and Ted R. Gurr. 2003. "Political Regime Characteristics and Transitions, Polity IV Dataset.".
URL: *http://www.cidcm.umd.edu/inscr/polity/*

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

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

Mason, T. David. 1999. "The Civil War in El Salvador: A Retrospective Analysis." *Laitin American* 34(3):179–196.

Mason, T. David and Dale A. Krane. 1989. "The Political Economy of Death Squads: Toward a Theory of the Impact of State-Sanctioned Terror." *International Studies Quarterly* 33(2):175–198.

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

Meernik, James D., Rosa Aloisi, Marsha Sowell and Angela Nichols. 2012. "The Impact of Human Rights Organizations on Naming and Shaming Campaigns." *Journal of Conflict Resolution* 56(2):233–256.

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.

Moore, Wil H. 2000. "The Repression of Dissent: A Substitution Model of Government Coercion." *Journal of Conflict Resolution* 44(1):107–127.

Moore, Will H. 1998. "Repression and Dissent: Substitution, Context, Timing." *American Journal of Political Science* 42(3):851–873.

Morgan, T. Clifton and Glenn Palmer. 2000. "A Model of Foreign Policy Substitutability: Selecting the Right Tools for the Job(s)." *Journal of Conflict Resolution* 44(1):11–32.

Morrow, James D. 2007. "When do states follow the laws of war?" *American Political Science Review* 101(3):559–572.

Most, Benjamin A. and Harvey Starr. 1984. "International Relations Theory, Foreign Policy Substitutability, and "Nice" Laws." *World Politics* 36(3):383–406.

Most, Benjamin A. and Harvey Starr. 1989. *Inquiry, Logic, and International Politics*. Columbia, SC: University of South Carolina Press.

Murdie, Amanda and David R. Davis. 2012. "Looking in the mirror: Comparing INGO networks across issue areas." *Review of International Organizations* 7(2):177–202.

Murdie, Amanda and Tavishi Bhasin. 2011. "Aiding and Abetting? Human Rights INGOs and Domestic Anti- Government Protest." *Journal of Conflict Resolution* 55(2):163–191.

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

Palmer, Alexis, Jocelyn Tomkinson, Charlene Phung, Nathan Ford, Michel Joffres, Kimberly Fernandes, Leilei Zeng, Viviane Lima, Julio Montaner, Gordon Guyatt and Edward Mills. 2009. "Does ratification of human-rights treaties have effects on population health?" *The Lancet* 373(9679):1987–1992.

Palmer, Glenn and Archana Bhandari. 2000. "The Investigation of Substitutability in Foreign Policy." *Journal of Conflict Resolution* 44(1):3–10.

Palmer, Glenn, Scott B. Wohlander and T. Clifton Morgan. 2002. "Give or Take: Foreign Aid and Foreign Policy Substitutability." *Journal of Peace Research* 39(1):5–26.

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

Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. Viking.

Piwowarczyk, Linda, Alejandro Moreno and Michael Grodin. 2000. "Health Care of Torture Survivors." *Journal of the American Medical Association* 284(4):539–541.

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

Poe, Steven C. 1990. "Human Rights and US Foreign Aid: A Review of Quantitative Studies and Suggestions for Future Research." *Human Rights Quarterly* 12(4):499–512.

Poe, Steven C. 1991. "U.S. Economic Aid Allocation: The Quest for Cumulation." *International Interactions* 16(4):295–316.

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., C. Neal Tate, Linda Camp Keith and Drew Noble Lanier. 2000. The Continuity of Suffering: Domestic Threat and Human Rights Abuse Across Time. In *Paths to State Repression: Human Rights and Contentious Politics in Comparative Perspective*, ed. Christian Davenport. Boulder: Roman Littlefield pp. 37–70.

Poe, Steven C., Dierdre Wendel-Blunt and Karl Ho. 1997. "Global Patterns in the Achievement of Women's Human Rights to Political and Economic Equality." *Human Rights Quarterly* 19(4):813–835.

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.

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

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

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

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

Powell, Emilia Justyna and Jefferey K. Staton. 2009. "Domestic Judicial Institutions and Human Rights Treaty Violation." *International Studies Quarterly* 53(1):149–174.

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Rasch, Georg. 1980. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press.

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.

Regan, Patrick M. 2000. "Substituting Policies during U.S. Interventions in Internal Conflicts: A Little of This, a Little of That." *Journal of Conflict Resolution* 44(1):90–106.

Richards, David L. and Ronald Gelleny. 2007*a*. "Good things to those who wait? National elections and government respect for human rights." *Journal of Peace Research* 44(4):505–523.

Richards, David L. and Ronald Gelleny. 2007*b*. "Women's status and economic globalization." *International Studies Quarterly* 51(4):855–876.

Richards, David, Ronald Gelleny and David Sacko. 2001. "Money With A Mean Streak? Foreign Economic Penetration and Government Respect for Human Rights in Developing Countries." *International Studies Quarterly* 45(2):219–239.

Risse, Thomas and Kathryn Sikkink. 1999. The Socialization of International Human Rights Norms into Domestic Practices: Introduction. In *The Power of Human Rights: Inter- national Norms and Domestic Change*, ed. Thomas Risse, Stephen Ropp and Kathryn Sikkink. Cambridge, MA: Cambridge University Press.

Rosas, Guillermo. 2009. "Dynamic latent trait models: An application to Latin American banking crises." *Electoral Studies* 28:375–387.

Rottman, Andy J., Christopher J. Fariss and Steven C. Poe. 2009. "The Path to Asylum in the US and the Determinants for Who gets in and Why." *International Migration Review* 43(1):3–33.

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*a*. *Death by Government: Genocide and Mass Murder in the Twentieth Century*. New Jersey: Transaction Publishers.

Rummel, Rudolph J. 1994*b*. "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* .

Salehyan, Idean and Cullen Hendrix. 2012. "Social Conflict in Africa Database Version 2.0." *Codebook and coding procedures* .

Saleyhan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull and Jennifer Williams. 2012. "Social Conflict in Africa: A New Database." *International Interactions* 38(4):503–511.

Schnakenberg, Keith E. and Christopher J. Fariss. 2012. "Dynamic Patterns of Human Rights Practices." *working paper* .

Schnakenberg, Keith E. and Christopher J. Fariss. Forthcoming. "Dynamic Patterns of Human Rights Practices." *Political Science Research and Methods* .

Schoultz, Lars. 1981. "US Foreign Policy and Human Rights Violations in Latin America: A Comparative Analysis of Foreign Aid Distributions." *Comparative Politics* 13(2):149–170.

Schrodt, Philip A. 2012. "Precedents, Progress, and Prospects in Political Event Data." *International Interactions* 38(4):546–569.

Settle, Jaime E., Robert M. Bond and Justin Levitt. 2011. "The Social Origins of Adult Political Behavior." *American Politics Research* 39(2):239–263.

Sijtsma, Klaas and Ivo W. Molenaar. 2002. *Introduction to nonparametric item response theory*. Sage: Thousand Oaks, CA.

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

Silove, Derrick. 1999. "The psychosocial effects of torture, mass human rights violations, and refugee trauma - Toward an integrated conceptual framework." *Journal of Nervous and Mental Disease* 187(4):200–207.

Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4):819–835.

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

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

Singer, J. David. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985." *International Interactions* 14(2):115–132.

Singer, J. David, Stuart Bremer and John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War, and Numbers*, ed. Bruce Russett. Beverly Hills: Sage pp. 19–48.

Singh, Jerome Amir, Michelle Govender and Edward J. Mills. 2007. "Health and human rights 2 - Do human rights matter to health?" *The Lancet* 370(9586):521–527.

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.

Smeulers, Alette. 2004. What Transforms Ordinary People into Gross Human Rights Violators? In *Understanding Human Rights Violations: New Systematic Studies*, ed. S. Carey and S. Poe. Aldershott: Ashgate pp. 239–253.

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.

Starr, Harvey. 2000. "Substitutability in Foreign Policy: Theoretically Central, Empirically Elusive." *Journal of Conflict Resolution* 44(1):128–138.

Stohl, Michael, David Carleton, George Lopez and Stephen Samuels. 1986. "State Violation of Human Rights: Issues and Problems of Measurement." *Human Rights Quarterly* 8(4):592–606.

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

Sundberg, Ralph. 2009. Revisiting One-sided Violence: A Global and Regional Analysis. In *States in Armed Conflict*, ed. Lotta Harbom and Ralph Sundberg. Uppsala: Universitetstryckeriet.

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.

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

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

United States Energy Information Administration. 2009.
URL: *http://www.eia.doe.gov/ (2008, December 5)*

van Schuur, Wijbrandt H. 2003. "Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory." *Political Analysis* 11(2):139–163.

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

Vreeland, James Raymond. 2008. "Political institutions and human rights: Why dictatorships enter into the United Nations Convention Against Torture." *International Organization* 62(165-101).

Wasserman, Stanley and Katherine Faust. 1994. *Social Networks Analysis: Methods and Applications*. Cambridge: Cambridge 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.

Wong, Wendy H. 2012. *Internal Affairs: How the Structure of NGOs Transforms Human Rights*. Ithaca, NY: Cornell University Press.

Wood, Reed M. 2008. "'A Hand upon the Throat of the Nation': Economic Sanctions and State Repression, 1976-2001A Hand upon the Throat of the Nation": Economic Sanctions and State Repression, 1976-2001." *International Studies Quarterly* 52(3):489–513.

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

World Bank. 2009. "Data and Statistics.".
URL: *http://go.worldbank.org/6HAYAHG8H0*

Zanger, Sabine C. 2000. "A Global Analysis of the Effect of Political Regime Changes on Life Integrity Violations, 1977-93." *Journal of Peace Research* 37(2):213–233.