

Supporting Information for “Torture, violent dissent and the palliative impact of democratic institutions”

David Armstrong^{*}, Christian Davenport[†], and Will H. Moore[‡]

^{*}University of Wisconsin, Milwaukee, USA, [†]Notre Dame University, South Bend, USA, and [‡]Florida State University, Tallahassee, USA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

This paper describes the research design, methods, variables, and data used to produce the results reported in “Torture, violent dissent and the palliative impact of democratic institutions”

Torture | Democracy | Measurement

This paper describes how we produced the results reported in the accompanying article. It is divided into three sections, the first of which describes our research design. In the next section we discuss the data we used to measure our variables, and describe the measurement models we used to construct empirical indicators for the latent variables we use. The third section describes the statistical models that we estimated and also presents additional results that provide further support for the inferences described in the article.

Research Design

Because our argument focuses on a comparison between states with democratic and autocratic institutions we need a sample with substantial variation across those institutions. We also require variation in both violent activity targeting the state (i.e., violent dissent) and state uses of torture. A cross-national sample is most appropriate, and though we do not restrict our hypotheses to the final quarter of the 20th Century, data availability limits our attention to that time frame. Our unit of observation, then, is the country-year, and we have a pooled cross-sectional time-series data structure. Below we discuss the estimation issues that arise.

Our research design has an important bias issue: our measure of torture is a biased under count of torture events. This is not our choice; under counts are inevitable in any empirical measure of an event that conspirators wish to hide from public view. Importantly, this bias is *not* a serious problem for our research design. As King, Koehane & Verba explain, truncation in the dependent variable produces biased inferences of the size of the effect of any independent variable on that dependent variable (2007, 130). In other words, our study will underestimate the substantive effect of regime type and threat on torture. More specifically, given that we develop a dichotomous indicator of torture (see below) our measure will misclassify states that successfully hide their use of torture as states that did not use torture. This type of measurement error is biased *against* our hypothesis and is thus the conservative sort of bias that one wants if one’s design or measures must contain bias.¹

Data & Measurement

We utilize two types of variables in our study: observed and latent. Observed variables are those that can be measured using a single indicator and for which we need only assume a small amount of random measurement error that can be adequately modeled using the error term in a regression equation. Latent variables, on the other hand, will exhibit both large random error *and* bias if they are measured with only one indicator. To address that problem we construct measurement models and estimate our latent variables using a vector of indicators. We first describe our data sources for the observed variables, and then turn our attention to the measurement model and data for our latent variables.

Observed Variables. To measure our dependent variable, we use data from the Cingranelli and Richards (CIRI) Human Rights database which codes the State Department and Amnesty International country re-

Reserved for Publication Footnotes

¹At a recent meeting Håvard Hegre expressed his belief that these data suffer from another, more tenacious, form of measurement error that we will call conditional bias: the extent to which the data have a downward bias is *not* constant across press freedom. Bollen and Goodman & Jinks assert that this is a serious problem in all human rights performance data, and Drakos & Gofas produce evidence to support the claim that the same measurement bias exists with respect to data that record the number of acts of terror (i.e., states that ensure freedom of the press will have systematically less under-reporting of terror events than states that restrict the press; 2007, 175-6, 2007). Interestingly, Li argues that—with respect to acts of terror—the conventional wisdom is misplaced, and produces statistical results consistent with his expectations (2007). Leaving the measurement bias in the terror data aside, measurement bias in our project this suggests that we have a systematic over-reporting of zeros (no torture) in countries that restrict the press. This is a different type of bias and requires a solution. We are unaware of work that reports and addresses this issue, and have set out to write a zero-inflated logit model. Zero-inflated models for event counts are well developed, but we are unfamiliar with efforts to extend this type of model to a logit setting. One of the authors of this study has recently developed, with others, an alpha version of such an estimator. Should we successfully develop this estimator, a future version of this study will report the results obtained when one uses that estimator. We should also note that Hathaway & Ho use a Bayesian factor analytic approach as a solution to generic measurement error (as opposed to the specific measurement bias considered here; 2007).

ports in 146 countries from 1980 to 1999.² The CIRI project defines torture as

the purposeful inflicting of extreme pain, whether mental or physical, by government officials or by private individuals at the instigation of government officials. Torture includes the use of physical and other force by police and prison guards that is cruel, inhuman, or degrading. Torture can be anything from simple beatings, to other practices such as rape or administering shock or electrocution as a means of getting information, or a forced confession (? , 12).

It thus includes acts of torture that mark the flesh as well as those that Rejali labels “stealth” (i.e., water, electricity, sleep deprivation, and stress positions), and is thus a valid indicator for our study (?). The CIRI project codes an ordinal variable based on the number of torture events reported in a given country-year, thus producing an under count of the real number. We use the three ordinal categories to create a dichotomy - torture (what CIRI codes as categories zero and one) and no torture (what CIRI codes as category 3). This does not eliminate the bias, as there are likely regimes coded as having zero incidents of torture that do engage in the activity. However, we feel that not trying to distinguish among gradations of torture mitigates the potential bias as much as possible. Whatever bias does remain works against finding support for our hypothesis and is thus conservative.

We selected CIRI rather than the Hathaway data for several reasons (?). First, Hathaway uses only U.S. State Department reports, which Poe et. al. suggest has a particular bias (?). CIRI uses both the State Department and Amnesty International data to code countries on human rights practices. Second, the Hathaway measure includes fewer countries (likely as a function of using the State Department Country Report). In particular, one of the countries excluded from Hathaway’s sample is the United States. Though this is not necessarily a fatal flaw, the fact that the behavior of the United States is motivating this investigation requires us to take lengths to ensure the U.S. is in the sample. Finally, the Hathaway sample includes fewer years (she starts at 1985 and CIRI starts at 1980). We prefer to leverage all of the data possible to test our argument.

To examine our hypotheses we also need to operationalize violent political dissent. To do so we use three different sources. We code a country as having violent dissent if there are any instances of guerrilla war [as defined and operationalized by Banks (?)] or civil war either as coded by the Correlates of War project (?), which imposes a 1000 battle death minimum over the course of the war, or the Peace Research Institute in Oslo (?) which imposes a much smaller 25 battle deaths/year minimum for civil violence to rise to the level of “civil war”. Since these events are relatively rare, and in this context providing estimates with any precision would be difficult, we create a variable we name violent dissent that is a dichotomous variable coded 1 when either of the violent conflict measures identified above exist and 0 when neither is present. We believe this this operationalization provides a small enough cutoff to show that even regimes facing low levels of internal violence are less likely to reap the palliative benefits of democracy.

We also include two control variables that have been found to effect the use of repressive strategies in a broad range of empirical studies: GDP per capita and Population³ and are also known to co-vary with our other variables of interest.⁴ To measure GDP/capita and population we use variables developed by Gleditsch, who took great pains to develop a complete dataset that took into consideration as much information as possible (?).

Latent Variables. To measure both the press and voice variables we employ a Bayesian procedure that estimates the latent variable that is thought to have produced the observed variables (??). The model has nice properties in that it does not require that all variables be observed for all cases, rather it uses existing information to find the most likely value of the latent variable (here, either voice or press) to have generated the data. Further, the model leverages the longitudinal nature of the data to temporally smooth the latent variable and as a result, decrease the variance of latent variable estimates. We discuss the details of the model in the section below that describes our statistical models, and here describe the observed indicators we use in those models.

Voice

We have identified five indicators that we use to estimate our latent measure for our voice concept. Dahl (1971) famously identified two dimensions of what he called polyarchy: competition and participation (?). We understand the former to be the extent to which entry costs to run for political office are the same across all citizens and groups of citizens. The closer to parity, the greater the level of competition. We conceptualize the second dimension to be the extent to which all adults have the right to vote. While none of the measures we use in this study are ideal operational indicators of these two dimensions, all of them measure at least one of the two dimensions, and each has been used by others to measure classical democracy (i.e., rule by the people).

The Political Regimes (aka ACLP) project adopts a classical conceptualization of democracy as a polity where competitive elections determine who holds the executive and legislative government offices (??).⁵ To be a democracy a regime must meet two criteria: first,

²This is available from <http://www.humanrightsdata.com/>. Significantly deviating from earlier efforts to measure physical integrity rights, which combined distinct strategies of repression (????), CIRI is an important project because it disaggregates human rights violations, individually identifying torture, extra-judicial killings, disappearance, political imprisonment, freedom of speech and press, freedom of religion, freedom of movement, freedom of assembly and association, political participation, worker’s rights, women’s political rights as well as women’s economic and social rights (?).

³Sambanis argues that population is likely robust in studies of civil war onset because studies do not normalize deaths by population (i.e., use a ‘per capita deaths’ threshold rather than ‘deaths’ threshold). Given that we use a dichotomous measure based on a counts of torture events rather than torture events per capita, we include population in our study (? , 822).

⁴While “we can omit a control variable if... [t]he omitted variable is uncorrelated with the included variable,” we must control for variables we know to be correlated with our primary variables of interest (? , 170). See, also, Ray’s work (? , 13-15). A number of studies have shown that both regime type and the presence of violent dissent are associated with per capita national income and population.

⁵We use the recently updated data from Cheibub & Gandhi, and invert their scale to create a variable, `aclpreg`, that has a value of zero for non-democracies and a value of one for democracies (?).

the state must select its executive via (indirect) election, and second opposition parties must have a realistic chance of winning elections.⁶ Note that this conceptualization focuses on competitiveness and does not take into consideration Dahl's second dimension: participation. Several of the other indicators we use explicitly measure both dimensions.

Like Przeworski, et al., Bernhard, Nordstrom & Reenock are interested in the durability of democratic rule (??). They produce a binary measure of democracy that explicitly builds on Dahl's (1971) minimalist definition. They begin by identifying cases that are coded as democratic by each of the following datasets: Jagers & Gurr's (1995) Polity III, Gasiorowski's (1996) Political Regime Change Dataset, Freedom House's Freedom in the World, and ACLP's Political Regime Data (p. 783) (???). They then excluded cases that failed to include at least 50% suffrage rates, those where fraud was judged to be sufficiently large to have changed the outcome, and those cases where extensive violence was judged to influence the outcome (pp. 784-5).⁷ The resulting variable, which we label `BNRdemoc`, is available online at: <http://www.personal.psu.edu/mhb5/data/data.htm>.

The third indicator we use to construct our latent indicator of voice is a combination of competition and participation originally collected by (?) and modified by (?). Competition is defined as the percentage of legislative seats held by all but the largest party and participation is the percentage of votes cast by enfranchised individuals.

$$G = \text{Competition} \times \log \left(\frac{\text{Participation}}{30} + 1 \right) \quad [1]$$

$$\tilde{G} = G * 0.98 + 0.01 \quad [2]$$

$$lG = \log \left(\frac{\tilde{G}}{1 - \tilde{G}} \right) \quad [3]$$

Finally, we use two components of the Polity project's measure of Democracy: Competitiveness of Participation and Competitiveness of Executive Recruitment (?). These are both measures of Dahl's competition dimension. Competitiveness of Participation, `parcomp`, "refers to the extent to which alternative preferences for policy and leadership can be pursued in the political arena" (? , 25). It has a six point scale ranging over the following values: 0, Unregulated (i.e., no competition); 1, Repressed; 2, Suppressed; 3, Factional; 4, Transitional; and 5, Competitive.

The second Polity indicator we use, Competitiveness of Executive Recruitment, `xrcomp`, measures the extent to which political advancement paths provide equal opportunities to non-incumbent individuals and groups. It has three values: 1, Selection (e.g., hereditary); 2, Dual/Transitional (e.g., dual executives, one selected via each); and 3, Competitive (e.g., multi-party elections; ?, 20).

Freedom of Expression

We construct our latent indicator of freedom of expression from three indicators. The first is Freedom House's measure of press freedom which ranks countries on a three-point scale of free, partly free and not free, based on the amount of editorial freedom within the country (?). The second, variable is Van Belle's (?)

five-point indicator of press freedom, which measures the extent to which the news media are "free" and "capable of functioning as an arena of political competition" (p. 138). There are four substantive categories—free, imperfectly free, restricted and controlled along with a "nonexistent or too limited to code" category, which we treat as missing. Finally, we use the three-point Freedom of Speech indicator from the Cingranelli and Richards Human Rights Database (?), which focuses on the extent of government censorship. The three categories indicate either complete government censorship, incomplete government censorship and no government censorship (i.e., a free press).

Statistical Methods

Recall that the hypothesis in the article leads us to expect that countries with both high voice and high press freedom will exhibit different lower predicted probabilities of "some" and "widespread" use of torture than countries with a high level of one variable, but a low level of the other, or low scores on both variables. The hypothesis also anticipates evidence that the positive joint impact of voice and press freedom on state's respect for the right to freedom from torture will be strongly reduced when a country experiences violent dissident attacks than when it does not. Figures 1 and 2 in the article depict the conditional predicted probabilities given no violent dissent and violent dissent for the jointly observed values of voice and press freedom. Ultimately, however, the evidence provided in Figures 1 and 2 of the article is suggestive: we did not report the results of formal hypothesis tests to show that the differences in the predicted probabilities across the quadrants in each Figure are statistically significant. In addition, we do not report in the article formal hypothesis tests to establish that the difference in those differences given no violent dissent and violent dissent are statistically significant. Below we report the results of formal hypothesis tests that support both of these necessary inferences, and thus substantiate the findings reported in the article.

In addition to reporting the results of the hypothesis tests we also provide a considerable amount of technical detail to support the replication of our analyses. We begin by describing the measurement models we used to estimate our latent variables, and show that diagnostic analyses support the use of those data. We then describe the predictive model that we use to generate the findings that we report in the article. We establish that the specification of the model we use is superior to two reasonable alternative specifications, and then we describe, and report the results of, the for-

⁶More specifically, there must be at least one successful transfer of power from one political party to a second. All regimes that fail to meet these two criteria are coded as dictatorships

⁷Note that much of the focus of their study is distinguishing among democracies. To make such distinctions they conceptualize along two dimensions: macro-institutional structure and party system. By macro-institutional structure they refer to executive-legislative configuration and, more specifically, the distinction between Presidential and Parliamentary systems (p. 778). They distinguish four types of party systems: predominant party; two party; moderate multiparty; and extreme multiparty. Following Colomer they conceptualize the interaction of the two dimensions along a single dimension anchored by Majoritarian and Pluralism (with quasi-majoritarian, mixed, and quasi-pluralism rounding out a five category scheme; see figure 1, p. 780; ?).

mal hypothesis tests required to support the findings we present in the article.

The Measurement Model. The indicators that we use to estimate our latent variables have three different levels of measurement. Two of the indicators are dichotomous (the ACLP and BNR measures of democracy); five of the measures are ordered categories (Competitiveness of executive recruitment and participation from Polity IV, and the three press measures); one of the measures is continuous (the polyarchy measure of Vanhanen and Gates). This makes traditional factor analysis an unappealing prospect as the relationships among the observed variables are not well-captured by a linear functional form (i.e., correlations). Thus, we use a Bayesian model that permits appropriate modeling of the relationships between the latent and observed variables.

Each measurement model takes roughly the same form, so we present the model in its most general form here and apply it to each set of indicators mentioned above. Let \tilde{Y} refer to the $N \times k$ matrix of observed variables (think of this as the five voice variables or the three press variables). The underlying measurement model, then suggests:

$$\tilde{Y} = f(\text{latent}) \quad [4]$$

Thus, we say that each observed variable is a stochastic function of its corresponding latent variable. The nature of $f(\cdot)$ is different for each different type of observed variable. The variables measured on a continuous scale will have a linear relationship to the latent variable, binary variables will be modeled with a logistic regression and ordinal variables will be modeled with an ordered logistic relationship. Symbolically, we can represent the relationship between the continuous variable and the latent as:

$$Y_i^{\text{continuous}} = \lambda_0 + \lambda_1 \text{Latent}_i + e_i \quad [5]$$

the relationship between the binary variables and the latent as:

$$Pr(Y_i^{\text{binary}} = 1) = \frac{1}{1 + \exp(-(\lambda_0 + \lambda_1 \text{Latent}_i))} \quad [6]$$

and the relationship between the ordinal variables and the latent as:

$$Pr(Y_i^{\text{ord}} = m) = \frac{1}{1 + \exp(-(\kappa_m - \lambda_1 \text{Latent}_i))} - \frac{1}{1 + \exp(-(\kappa_{m-1} - \lambda_1 \text{Latent}_i))} \quad [7]$$

There are two different latent variables here—press and voice.

Priors

The use of Bayesian MCMC techniques to estimate the model requires the specification of prior distributions for all model parameters. Parameters here refer to both the coefficients relating the latent variables to their corresponding observed variables as well as the latent variable points themselves. For identification

purposes, we set the coefficients relating both the Vanhanen/Gates variable and the CIRI free speech variable to their respective latent variables to 1. This imposes both a directionality of the latent variable and a scale. The other coefficients and intercepts (λ parameters) were given standard normal priors. The latent variables were given random walk priors such that in the first period (i.e., the first year the observation is in the dataset), they were drawn from a normal distribution with zero mean and variance of σ_p^2 for press and σ_v^2 for voice. In subsequent periods, the latent variables are drawn from normal distributions with mean of press_{t-1} or voice_{t-1} and variances of ν_p^2 and ν_v^2 , respectively. Thus the σ parameters set the scale of the latent variables and the ν parameters indicate how big changes are over time. Both the σ and ν parameters are given inverse gamma priors with both shape and rate equal to one. The κ terms are given censored normal priors to preserve the posterior ordering of the ordered-logit cutpoints.⁸

Measurement Results

There are two important sets of results produced by the measurement model. These results indicate the extent to which the observed variables are “good” indicators of the latent construct. If the variables are all (or mostly) good indicators of the construct, then this is evidence that the latent variable we use in a predictive model is capturing the phenomenon of interest. If this is not the case, then we can have no confidence that the latent variable is measure what we think it is.

The first of these results is the set of coefficients (λ parameters from above). These describe how well the latent variable relates to the observed indicators. These give us a sense of the reliability of each indicator. Figure ?? shows these coefficients in the first column. Note, that all of the λ parameters are statistically different from zero (as their 95% credible intervals do not include zero). This tells us that the latent variable has some explanatory power with respect to each observed variable.

⁸These models were estimated in WinBUGS version 1.4.3 (?) on a PC running Windows XP version 2002, SP3. The models were run for 10000 burn-in iterations and then 2500 iterations from each of two Markov chains were monitored for all model parameters. The results below are based on these 5000 (2500×2) iterations.

Variable	λ
Polyarchy (Gates)	1.00 (1.00, 1.00)
ACLP	0.27 (0.23, 0.30)
BNR	0.22 (0.18, 0.26)
XRCOMP (Polity IV)	1.60 (1.49, 1.72)
PARCOMP (Polity IV)	2.98 (2.67, 3.29)
Speech (CIRI)	1.00 (1.00, 1.00)
Press Freedom (FH)	-1.22 (-1.40, -1.01)
Press Freedom (VB)	-5.60 (-6.32, -4.89)

*NB: Main entries are posterior medians
95% credible intervals in parentheses*

Fig. 1. Measurement Model Results: λ

Perhaps more enlightening are the results in Figure ??, the proportional reduction in error and Heron's expected proportional reduction in error (?).⁹ The *PRE* is essentially like an R^2 in the linear model, in fact R^2 is, itself a *PRE* measure. The *ePRE* measure is a modification of the *PRE* that acknowledges the fact that a model correctly predicting a one with probability 0.9 is better than a model that correctly predicts the same one with a probability of 0.51.

The *PRE* and *ePRE* statistics tell us whether the observed variables are good indicators of the latent construct. Looking down the columns of the table, the numbers generally appear to be relatively high. That is to say, the latent variable does a good job at explaining the observed indicators. This provides evidence that the latent variables are, in fact, measuring what we think they are measuring. The only outlier here is the Bernhard et. al. variable, which sees a very small proportional reduction in error, though a reasonably good *ePRE*). One potential reason for this is that the overwhelming majority of the cases are non-democracies (71% in our data), which leaves comparatively little for the voice variable to explain (a model with no variables could correctly classify 71% of the observations). The fact that the *ePRE* is much higher than the *PRE* means that the latent variable model is doing a better job of predicting those observations already correctly classified by the null model. This suggests that even though the Bernhard et. al. variable is not the most reliable indicator, the latent variable has some predictive power on this variable.

As for the press variable, Van Belle's indicator seems to be the most reliable with *PRE* and *ePRE* values almost twice as big as the other two variables. That, said, the statistics indicate that both the Freedom House and CIRI variables are sufficiently well-explained by the press freedom latent variable to suggest that they are reliable indicators of the latent construct.

Variable	RE	ePRE
Polyarchy (Gates)	0.88 (0.87, 0.90)	NA
ACLP	0.76 (0.65, 0.80)	0.23 (0.20, 0.26)
BNR	0.02 (0.00, 0.09)	0.13 (0.10, 0.16)
XRCOMP (Polity IV)	0.56 (0.50, 0.59)	0.45 (0.41, 0.47)
PARCOMP (Polity IV)	0.68 (0.50, 0.73)	0.59 (0.49, 0.62)
Speech (CIRI)	0.41 (0.38, 0.43)	0.35 (0.33, 0.36)
Press Freedom (FH)	0.48 (0.44, 0.52)	0.40 (0.38, 0.42)
Press Freedom (VB)	0.86 (0.76, 0.89)	0.81 (0.75, 0.83)

*NB: Main entries are posterior medians
95% credible intervals in parentheses*

Fig. 2. Measurement Model Results: PRE and ePRE

We are most interested in the estimates of the latent variables themselves. As these are Bayesian models, each voice and press estimate (i.e., a country year) has a distribution. These will provide the main variables of interest in the regression models along with violent dissent. To give a sense of how these vary over time and across countries, see Figures ??-??. The posterior means of the voice variable range from -2.85 to 5.31 with a mean of 0.53. The posterior means of the press variable range from -5.71 to 6.25 with a mean of -0.02. In the figure, each variable was centered around its grand mean, making the mean of both centered variables zero.

Some scholars might try to argue that the correlation between these two variables is high enough (at 0.81) to suggest that there is really only one concept of interest - democracy of which press and voice are both indicators. While press and voice exhibit within-country co-variation, there are plenty of instances where one increases and the other either makes no change or decreases - indicating that there is no

⁹Let's take the binary case as an example. If we define *PCP* as the percentage correctly predicted (i.e., the percentage of ones for which the predicted probability is greater than 0.5 and the percentage of zeros for which the predicted probability is less than 0.5), and *PMC* as the percentage in the modal category (i.e., the null *PCP*), then

$$PRE = \frac{PCP - PMC}{1 - PMC} \quad [8]$$

The expected *PRE* is defined similarly, however the expected *PCP* is defined as:

$$ePCP = \sum_{y=1} \widehat{PR}(y=1) + \sum_{y=0} (1 - \widehat{PR}(y=1)) \quad [9]$$

We can then define *ePMC* (the expected Percentage in the Modal Category) as:

$$ePMC = \sum_{y=1} \widehat{PR}_0(y=1) + \sum_{y=0} (1 - \widehat{PR}_0(y=1)) \quad [10]$$

where $\widehat{PR}_0(y=1)$ is the predicted probability that $y=1$ under the null model, which is just the marginal probability of $y=1$ in the observed y . Then,

$$ePRE = \frac{ePCP - ePMC}{N - ePMC} \quad [11]$$

This extends straightforwardly to polychotomous variables.

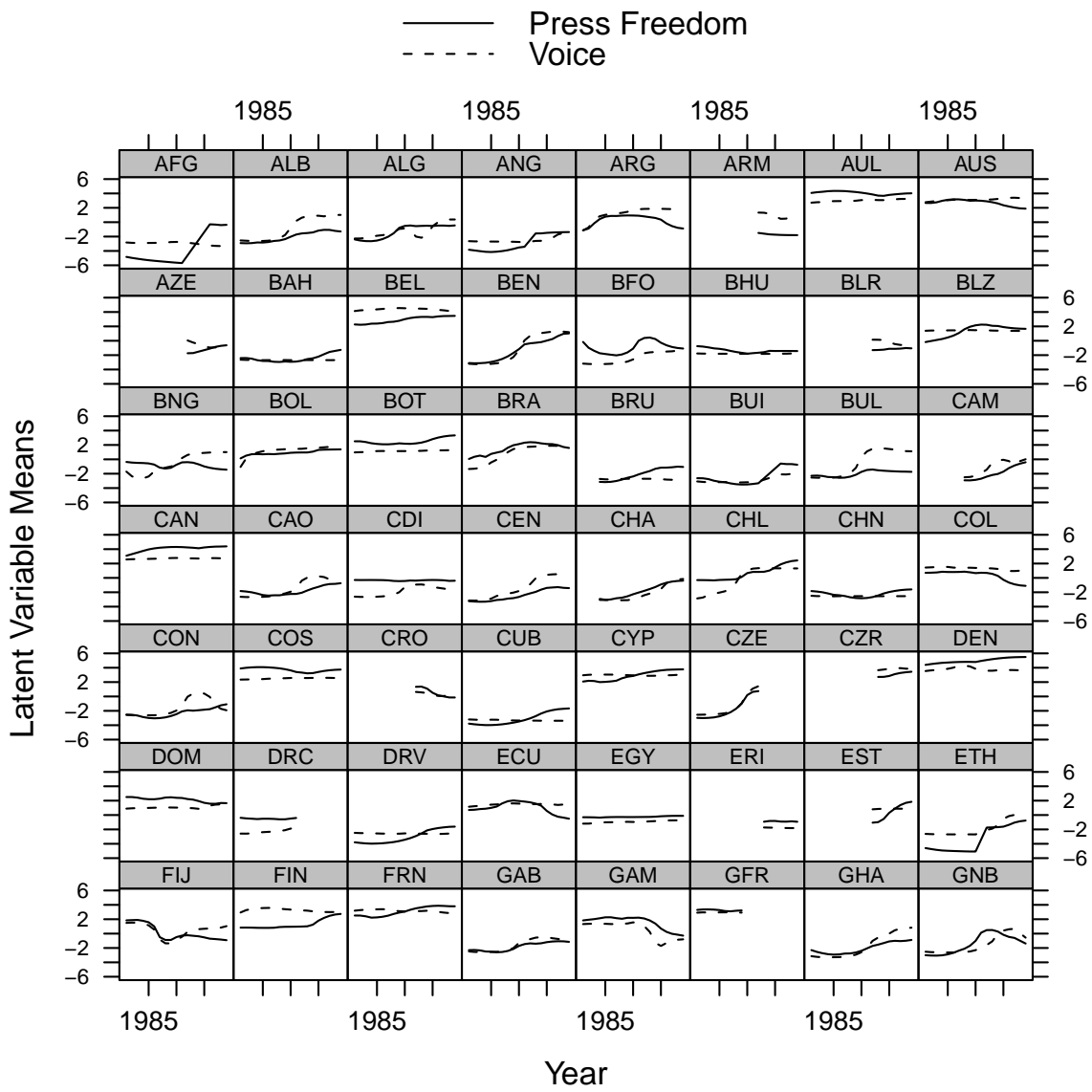


Fig. 3. Voice and Press over Time by Country I

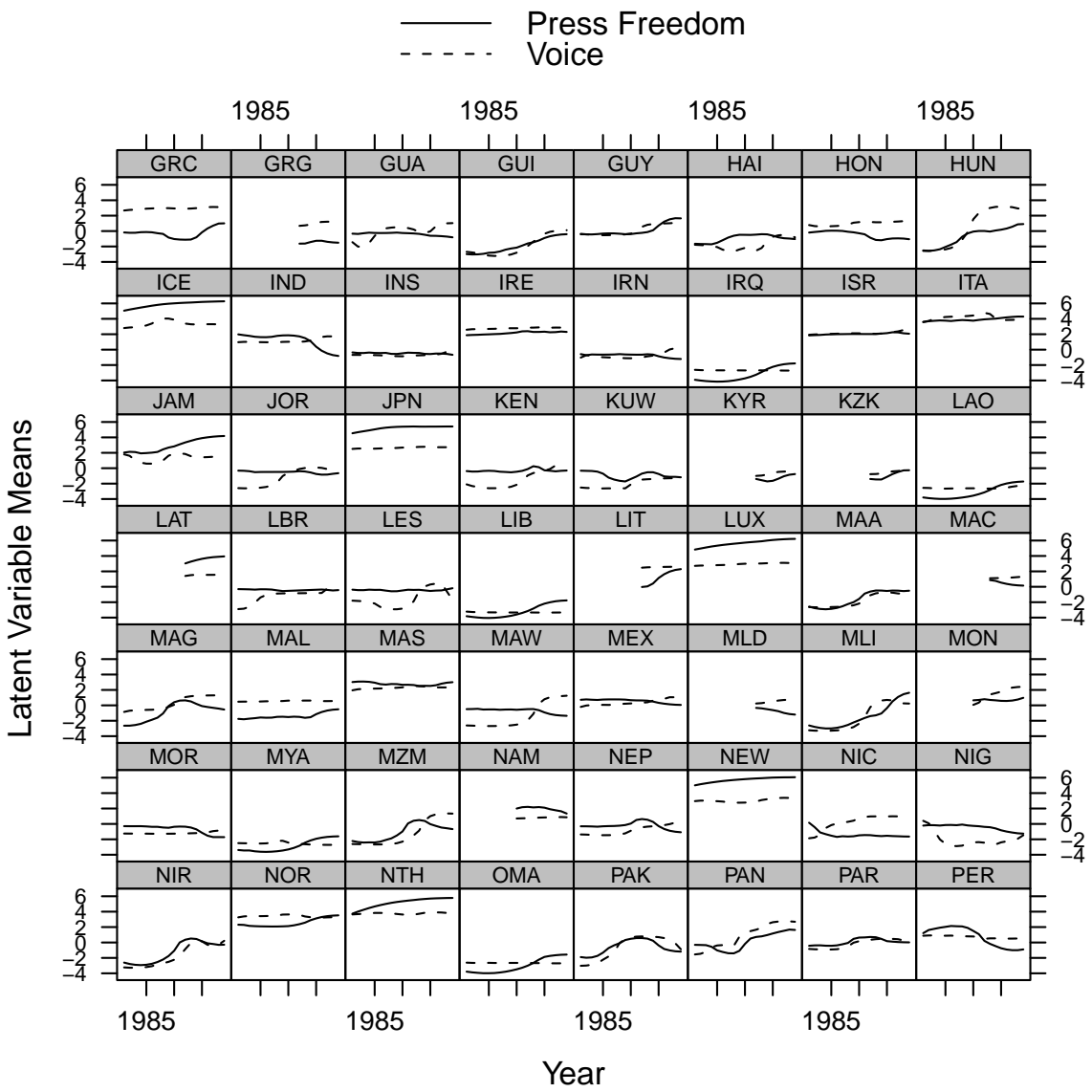


Fig. 4. Voice and Press over Time by Country II

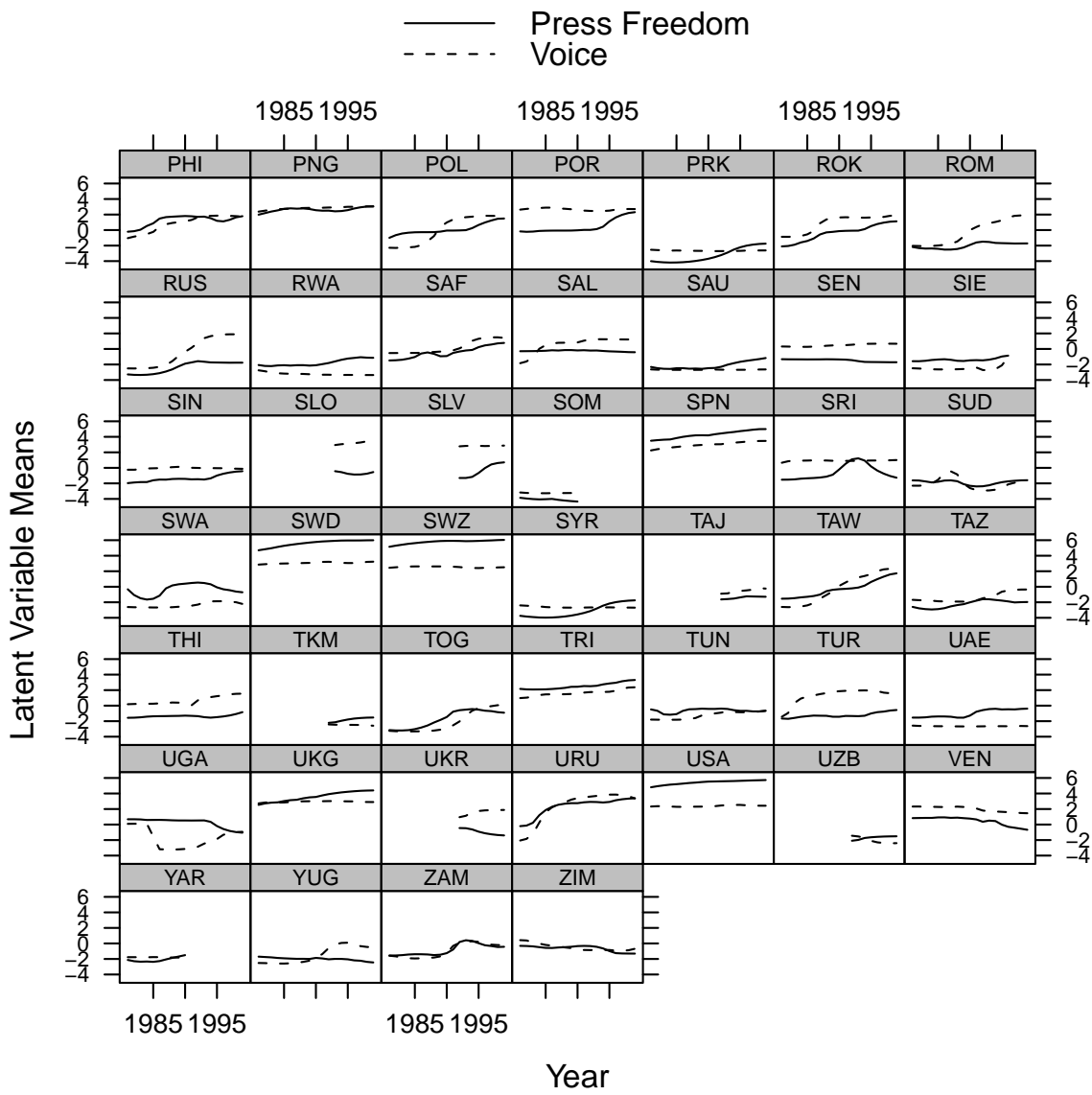


Fig. 5. Voice and Press over Time by Country III

necessary relationship here. 36% of the within-country correlations between voice and press are negative and 23% of the positive correlations are less than 0.5. Thus, we argue this is sufficient evidence to suggest these are related, but distinct concepts.

The Predictive Model. For the predictive model we estimate a Generalized Additive Markov Transition models. This uses the Generalized Additive Model which allows the relationship between the response and covariates to be represented by an arbitrary smooth function (??). This function is estimated using generalized cross-validation to prevent overfitting of the data. We use this statistical model in the context of a binary Markov transition model (?). Thus, we are estimating the probability that a country does not use torture given that it was not torturing in the previous period; we are estimating the probability of transition away from respecting away from respect for this basic human right. Specifically, we are estimating:

$$\begin{aligned} & \log \left(\frac{Pr(\text{No Torture}_t | \text{No Torture}_{t-1})}{1 - Pr(\text{No Torture}_t | \text{No Torture}_{t-1})} \right) \quad [12] \\ & = f(\text{voice, press} | \text{dissent}) + f(\text{voice, press} | \text{no dissent}) \\ & + \beta_1 \text{GDP/capita} + \beta_2 \log(\text{Population}) \end{aligned}$$

Note that the arbitrary smooth functions of voice and press are conditional on violent dissent, thus allowing the relationship between voice, press and torture to be different in the presence and absence of violent dissent. These smooth functions are not easily framed in terms of a small set of coefficients (e.g., a cubic polynomial), so we do not attempt to represent them as such. Rather, we simply show the χ^2 test and its corresponding p -value to indicate whether including the variables as an arbitrary smooth function is statistically better than not including them at all.¹⁰

Using the results of the measurement model as covariates in the regression presents an additional hurdle. The Bayesian model acknowledges that the latent variable estimates are estimates with accompanying uncertainty (i.e., they have distributions). By using the posterior means of the latent variable scores as observed data (i.e., not accounting for their distributions), we may get over-confident estimates of the effects of the variables of interest. As such, we run the model above 5000 times, once on the latent variable scores from each iteration of the Markov chain. Each time through, we record the predicted values (holding population and GDP/capita constant at their median values) for values across the range of voice, press and violent dissent, the degrees of freedom used for smooth terms and the results of a test of the model with arbitrary smooth functions of voice and press to one that includes an interaction between voice, press and violent dissent, though all in a linear fashion.

The results are convincing. More than 99% of the 5000 models showed that the generalized additive model was better than the model that did not allow for an arbitrary smooth relationship between voice, press and torture conditional on violent dissent. This provides evidence that we are not overfitting the data.¹¹ On average, the relationship between voice, press and torture when there is no violent dissent used around 5 degrees of freedom whereas the relationship between voice, press and torture in the presence of violent dis-

sent used 2 degrees of freedom (indicating no interaction at all). The relationship between voice, press and torture in the presence of violent dissent was statistically different from zero (i.e., was not a flat plane) in only 5 of the 5,000 models. In all 5,000 models the relationship between voice, press and torture in the absence of violent dissent was significant. Together, all of this evidence suggests that there is a significant, non-linear relationship between voice, press and torture when there is no violent dissent and there is essentially no relationship between voice, press and violent dissent in the presence of violent dissent. You can see these probabilities in Figure ??.

Fig. 6. Relationship of Voice, Press and Dissent to Torture

¹⁰The models were run in **R** (64 bit) version 2.10.1 (?), using the **mgcv** library version 1.6-1 (?) on a MacPro with OSX version 10.5.8.

¹¹Further to this point, in each model, the maximum degrees of freedom the smoother was able to use was fixed at 10 for both terms. While this, as many choices in statistics was arbitrary, it was found to be robust.

References

- King G, Keohane R, Verba S (1994) *Designing Social Inquiry* (Princeton University Press, Princeton).
- Bollen KA (1992) in *Human Rights and Statistics: Setting the Record Straight*, eds Jabine T, Claude R (University of Pennsylvania Press, Philadelphia)
- Goodman R, Jinks D (2003) Measuring the effects of human rights treaties. *European Journal of International Law* 14:171–184.
- Drakos K, Gofas A (2006) The devil you know but are afraid to face: Underreporting bias and its distorting effects on the study of terrorism. *Journal of Conflict Resolution* 50:714–735.
- Li Q (2005) Does democracy promote or reduce transnational terrorist incidents? *Journal of Conflict Resolution* 49:278–297.
- Hathaway OA, Ho DE (2004) Characterizing measurement error in human rights. Paper presented at the annual meeting of the American Political Science Association.
- Cingranelli DL, Richards DL (1999) Measuring the level, pattern and sequence of government respect for physical integrity rights. *International Studies Quarterly* 43:407–418.
- Henderson C (1991) Conditions affecting the use of political repression. *Journal of Conflict Resolution* 35:120–142.
- Mitchell NJ, McCormick JM (1988) Economic and political explanations of human rights violations. *World Politics* 40:476–498.
- Poe S, Tate CN (1994) Repression of personal integrity rights in the 1980's: A global analysis. *American Political Science Review* 88:853–872.
- Cingranelli DL, Richards DL (2004) The cingranelli-richards (ciri) human rights dataset. <http://www.humanrightsdata.org>, accessed 12-1-2004.
- Cingranelli DL, Richards DL (2004) The cingranelli richards (ciri) human rights database coding manual. available online at: <http://ciri.binghamton.edu/documentation.asp>.
- Rejali D (2007) *Torture and Democracy* (Princeton University Press, Princeton, New Jersey).
- Hathaway OA (2002) Do human rights treaties make a difference? *Yale Law Journal* 111:1935–2042.
- Poe SC, Vazquez T, Zanger S (2001) How are these pictures different: Assessing the biases in the u.s. state department's country reports on human rights practices. *Human Rights Quarterly* 23:650–677.
- Banks AS (2005) *Cross-National Time-Series Data Archive* (Databanks, Intl., Jerusalem, Israel).
- Sarkees MR (2000) The correlates of war data on war: An update to 1997. *Conflict Management and Peace Science* 18:123–144.
- Gleditsch NP, Wallensteen P, Eriksson M, Sollenberg M, Strand H (2002) Armed conflict 1946 - 2001: A new dataset. *Journal of Peace Research* 39:615–637.
- Sambanis N (2004) What is civil war? conceptual and empirical complexities of an operational definition. *Journal of Conflict Resolution* 38:814–858.
- Ray JL (2003) Explaining interstate conflict and war: What should be controlled for? *Conflict Management and Peace Science* 20:1–31.
- Gleditsch K (2002) Expanded trade and gdp data. *Journal of Conflict Resolution* 46:712–24.
- Armstrong DA (2009) Measuring the democracy-repression nexus. *Electoral Studies* 28(3):403–412.
- Jackman S (2009) *Bayesian Analysis for the Social Sciences* (John Wiley & Sons, Ltd., Chichester, UK).
- Dahl R (1971) *Polyarchy: Participation and Opposition* (Yale University Press, New Haven).
- Alvarez M, Cheibub J, Limongi F, Przeworski A (1996) Classifying political regimes. *Studies in Comparative International Development* 31:1–37.
- Przeworski A, Alvarez M, Cheibub JA, Limongi F (2000) *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990* (Cambridge University Press, New York).
- Cheibub JA, Gandhi J (2004) Classifying political regimes: A six fold classification of democracies and dictatorships. Dataset, Yale University.
- Przeworski A, Alvarez M, Cheibub JA, Limongi F (1996) What makes democracies endure? *Journal of Democracy* 7:31–55.
- Bernhard M, Nordstrom T, Reenock C (2001) Economic performance, institutional intermediation and democratic breakdown. *Journal of Politics* 63:775–803.
- Jagers K, Gurr T (1995) Transitions to democracy: Tracking the third wave with polity iii indicators of democracy and autocracy. *Journal of Peace Research* 32:469–482.
- Gasiorowski MJ (1996) An overview of the political regime change dataset. *Comparative Political Studies* 29:469–483.
- Freedom House (2005) *Freedom in the World* (Freedom House, New York).
- Colomer JM (1995) Strategies and outcomes in eastern europe. *Journal of Democracy* 6:74–85.
- Vanhanen T (2000) A new dataset for measuring democracy, 1810-1998. *Journal of Peace Research* 37:251–265.
- Gates S, Hegre H, Jones MP, Strand H (2006) Institutional inconsistency and political instability: The duration of polities. *American Journal of Political Science* 50:893–908.
- Marshall M, Jagers K (2005) *Polity IV Project: Dataset Users Manual* (Center for International Conflict Data Management) available online at: <http://www.cidcm.umd.edu/polity/>.
- Freedom House (2006) *Freedom of the Press 2006: A Global Survey of Media Independence* (Rowman & Littlefield, Lanham, MD)
- Van Belle DA (2000) *Press Freedom and Global Politics* (Praeger Publishers, Westport, CT).
- Cingranelli DL, Richards DL (2004) The cingranelli richards (ciri) human rights database coding manual. available online at: <http://ciri.binghamton.edu/documentation.asp>.
- Spiegelhalter D, Thomas A, Best N, Lunn D (2007) *WinBUGS User Manual, version 1.4.3* (Cambridge, UK).
- Herron MC (1999) Postestimation uncertainty in limited dependent variable models. *Political Analysis* 8:83–98.
- Hastie T, Tibshirani R (1990) *Generalized Additive Models* (Chapman and Hall, New York).
- Wood S (2006) *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC).
- Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data* (Oxford University Press, Oxford).

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing* (R Foun-

dation for Statistical Computing, Vienna, Austria)
ISBN 3-900051-07-0.