# 1    Online Appendix

## 1.1    Control Variables and Descriptive Statistics

In the manuscript, I utilize several control variables in the models to account for alternative explanations of government torture. First, torture tactics are often used simultaneously. Following Conrad, Hill and Moore (2018), I include counts of the number of scarring *(Scarring)*, clean *(Clean)*, and unstated *(Unstated)* torture in the models. For example, in the model predicting scarring torture, I include counts of unstated and clean torture. Second, torture is more likely to occur when states face threats to their control, including violent dissent (Conrad and Moore 2010). As a result, I include a measure of civil conflict *(Civil War)* from Nils Gleditsch and Strand (2002).

Third, several domestic institutions are found to be related to domestic mobilization and torture. Judicial independence *(Judicial Independence)* is found to be positively related to torture tactics, specifically clean torture (Conrad, Hill and Moore 2018), and I include a measure from Coppedge et al. (2016) capturing high court independence. I also include a binary variable capturing whether the state has adopted a national human rights institution (*NHRI*), as Welch (2017) finds that NHRI presence is associated with state torture practices. Data on NHRIs is taken from the Organizational NHRI Project Dataset for the year the NHRI was formally occupied (Conrad et al. 2013). I also control for freedom of *Speech* (Cingranelli, Richards and Clay 2014) because free speech protections allow individuals to speak out about rights abuses.

In addition, I control for the extent to which the state is embedded in the international human rights regime *(Embeddedness)*, taken from Fariss (2018), as socialization increases the likelihood of mobilization and better rights practices. Economic development and country wealth are also found to be associated with respect for physical integrity rights, including torture (Poe and Tate 1994). As a result, I also include variables accounting for country wealth, (logged GDP per capita) *(GDP)* and (logged) population *(Population)*.

The ITT data codes *allegations* of torture made by AI and allegations of torture do not represent the amount of torture (or violations) occurring within a state. As a result, scholars interested in

drawing inferences about the actual rights behavior of states must account for the strategic process by which AI generates allegations. Following Conrad, Hill and Moore (2018), I utilize several control variables that account for the likelihood that AI makes allegations. First, I utilize a count of human rights organizations *(HRO)* with a secretariat office in the state utilizing data from (Murdie and Bhasin 2011). In addition, I include a variable from the ITT dataset accounting for whether or not AI published a statement that AI, or another INGO had difficulty gaining access to detained individuals and victims *(Restricted Access)* (Conrad and Moore 2010).

Finally, AI may rely on beliefs about the state's human rights practices in deciding to make an allegation (learn about and report an allegation) (Conrad, Hill and Moore 2018). As a result, I include a variable accounting for lagged physical integrity rights, using the Fariss (2014) latent human rights protection scores. In addition to accounting for AI's beliefs, Conrad, Hill and Moore (2018) argue that accounting for lagged respect for physical integrity rights "helps us model any bureaucratic inertia in the reporting and serves as a proxy measure for the size and strength of the grass roots network in each country that AI taps to obtain information" (10). Table 1 displays descriptive statistics for the variables included in the models.

[Table 1 about here.]

## 1.2   Statistical Tables

I report statistical tables from the results displayed in the manuscript in Table 2 below. Although the coefficient plots in the main manuscript report coefficient estimates and 90 percent confidence intervals, Table 2 displays coefficient estimates and standard errors. Reported results are significant at the p < .05 level for the scarring torture model and significant at the p < .10 level for the formal complaint model.

[Table 2 about here.]

## 1.3 Aggregate Torture Model Results

To illustrate the importance of disaggregating the torture dependent variable into types of torture tactics (e.g. scarring and stealth), I estimate a model using the Cingranelli, Richards and Clay (2014) torture data as the dependent variable. The Cingranelli, Richards and Clay (2014) torture variable is an ordinal variable in which a zero indicates that torture is practiced frequently, a one indicates that torture is practiced occasionally, and a two indicates that torture is not practiced. Given the nature of the dependent variable, I estimate an ordered logistic regression model. The model results are displayed in Table 3. The results show the absence of a significant relationship between adverse Inter-American Court judgments or adverse Inter-American Commission decisions and the level government torture. These results provide evidence in favor of considering state adjustments in torture tactics in response to international institutions. By only looking at overall levels of torture, scholars may miss more nuanced (but important) changes in state repression.

[Table 3 about here.]

## 1.4 Random and Fixed Effects Negative Binomial Regression Results

The heterogeneity across states indicates that each country's baseline probability of human rights policy change in response to adverse decisions from the Inter-American Human Rights System (Court or Commission) is likely not the same, even accounting for the influence of various control variables in the model. As a robustness check, I estimate models accounting for unobserved heterogeneity using multilevel modeling techniques (here, incorporating varying intercepts) (Gelman and Hill 2007:259). A multilevel model removes the restriction that the intercepts are constant across individual cases, and treats the cross-sectional deviations from the common intercept as random, rather than estimable. Results from random effects negative binomial regression models are displayed in Table 4 and the main variables of interest are robust to the inclusion of random effects.

[Table 4 about here.]

Given the short time-frame and therefore, the limited number of observations, country- and year-fixed effects grind up degrees of freedom and discard cross-sectional variation, making the model inefficient. The random effects framework incorporates cross-sectional and within-unit variation and is more efficient as it only estimates a variance parameter rather than N (or N-1) additional parameters. Models incorporating country fixed effects are reported in Table 5 and although all coefficients remain in the expected direction, because the models have relatively few observations, the standard errors grow beyond statistical significance in the models. That is, they only achieve statistical significance at the $p < .10$ level using a one-tailed significance test. Table 6 reports results from models incorporating year fixed effects. Given the limited time-frame, the results are robust to the inclusion of year fixed effects at the $p < .10$ level.

[Tables 5 and 6 about here.]

## 1.5 Standard Errors Clustered on Country

While all models are estimated with Huber-White standard errors, I also estimated the models with standard errors clustered on country. The cluster-adjusted standard error takes into account within-cluster correlation. Results are robust to the inclusion of standard errors clustered on country.

[Table 7 about here.]

## 1.6 Alternative Explanations of Compliance with International Law

In addition to the role of domestic politics and domestic mobilization efforts, scholars also point to the importance of selection and management issues in explaining the level of compliance with international agreements. First, some scholars argue that high compliance is the result of a selection effect, in which states that enter into international agreements are those that are compliant with the treaty *ex ante*, leading to little change in behavior following treaty commitment (Downs, Rocke and Barsoom 1996). There are several reasons that this argument is not helpful in explaining the

4

influence of the Inter-American Human Rights System on state behavior and I examine selection in more detail below.

Second, others argue that managerial issues are a primary cause of noncompliance with international treaties (Chayes and Chayes 1993). Management problems include ambiguity in treaty language, lack of technical capacity to implement the treaty, or treaty provisions that are too ambitious to implement, among numerous other issues. Capacity limitations may explain some variation in state response to international human rights law, as states with greater resource capacity are in a better position to implement Commission recommendations and Court judgments. As a result, I control for logged GDP per capita in the main models. I also consider the extent to which the Court and Commission may render adverse judgments and decisions based on expectations about capacity below ("Examining Selection"). However, ambiguity in treaty provisions is less likely to create problems that make compliance with the Inter-American Commission and Court difficult. Both the Inter-American Commission and Court give the state a checklist of specific steps that the state must take to bring itself into compliance (Hawkins and Jacoby 2010). While treaty provisions may create ambiguity by leaving room open for interpretation, the Commission and Court *interpret* the American Convention on Human Rights and give states a list of steps necessary for compliance based on their interpretation. Given these arguments, in the article, I focus on the role of domestic politics and domestic mobilization in explaining state response to the Inter-American Court and Commission.

## 1.7   Sample of All Commission Members

In the article, I estimate models to examine the influence of adverse Inter-American Commission decisions on the number of torture allegations and the number of formal complaints for all states under the jurisdiction of the Inter-American Court of Human Rights for the years 1995-2005, which includes 21 states. The 21 states included in the analysis are Argentina, Barbados, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Uruguay, and Venezuela. These

21 states are included in the sample used to estimate the models because they are states under the jurisdiction of the Inter-American Court as well under the jurisdiction of the Inter-American Commission. As such, there is a non-zero probability all states included in the sample receive adverse Commission decisions and Court judgments.

As a robustness check, I estimate models including all states subject to the jurisdiction of the Inter-American Commission. In total, 35 states that are members of the Commission. This includes 14 additional states not included in the analyses in the manuscript. The 13 additional Inter-American Commission member states include: Antigua and Barbuda, Bahamas, Belize, Canada, Cuba, Dominica, Grenada, Guyana, Jamaica, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and Grenadines, Trinidad and Tobago, and the United States. I continue to exclude Trinidad and Tobago from the analysis because Trinidad and Tobago denounced the American Convention on Human Rights in 1998, an indication that it does not intend to comply or implement Commission decisions. In order to take into account the Commission's potential to influence all Commission members, states that have not accepted the Court's jurisdiction are included in the sample and I estimate a model including all 34 member states (excluding Trinidad and Tobago). In this sample, there were 118 adverse Commission decisions involving violations of article 5 of the American Convention on Human Rights, compared to 103 violations in the sample in the main manuscript.

Results (coefficient estimates and standard errors) are reported in Table 8 and they are largely robust to this larger sample. Adverse Commission decisions are not significantly related to the scarring and clean torture, though adverse Court judgments continue to be negatively and significantly related to scarring torture and insignificantly related to clean torture. Adverse Commission decisions remain positively related to formal complaints, but are only significant using a slightly less stringent evaluation of statistical significance, a one-tailed test of significance. Arguably, for states that are only subject to the Inter-American Commission, potential litigants may be less likely to see the Commission's decisions as a signal of legitimacy for their rights-related complaints. After all, their governments have not fully committed to the Inter-American Human Rights system. Notably, when I estimate a model using a sample of only Commission members that have not

accepted the Inter-American Court's jurisdiction (13 states), adverse Commission decisions have little influence on the likelihood of filing formal complaints.

[Table 8 about here.]

## 1.8  Examining Selection

Selection problems represent a serious concern for research examining the influence of international human rights law on state behavior (Hill 2010; Lupu 2013). One way a selection problem may manifest is due to state self-selection into treaties, which makes it difficult to determine whether international treaties have a causal effect on the behavior of the state. That is, changes in state behavior may not be caused by a treaty, but by the fact that states would have taken such actions anyway. Selection may be a problem for analyzing international human rights law because states that commit to international human rights law may be different from states that do not; these differences may explain state decisions to select into an international treaty and their decision to repress citizens, which "makes it difficult to separate the effect of the treaty from the effect of the institutional features that led them to ratify" (Hill 2010). Selection would be a problem for the analyses in this manuscript if states that have submitted to the jurisdiction of the Inter-American Courts of Human Rights are different from states that have not in important ways. There are 35 members of the Organization of American States (OAS), but currently only 20 members accept the jurisdiction of the Court. The members that have not accepted the jurisdiction of the Inter-American Court are quite diverse in terms of human rights practices, democratic consolidation, and size. For example, the average level of respect for rights for OAS member states that have not submitted to the jurisdiction of the Court is 6.2, while the average respect for rights for only those countries under the jurisdiction of the Court is 4.25. This indicates that countries that have ceded sovereignty to the Inter-American Court of Human Rights are not a subset of states that are in compliance *ex ante*.

A second way the selection problem may manifest is rooted in the idea that adverse Commission decisions and Court judgments are non-random. That is, adverse decisions from the Inter-

American Commission and adverse judgments from the Inter-American Court may be rendered where an expectation of compliance is more likely in order to protect the legitimacy of the Commission or Court. For example, adverse Commission decisions and Court judgments may be made more often in states with relatively higher levels of respect for rights, stronger democratic institutions, and greater capacity to implement judgments. Figures 1 and 2 provide descriptive evidence that states receiving adverse Commission decisions and Court judgments are not better rights-respecters than non-adverse decision or judgment recipients, which suggests that the Commission and Court are not making decisions based on where they expect a greater likelihood of compliance or where states are already in compliance. In fact, the Commission and Court are more likely to make adverse decisions in states with relatively worse human rights practices.

[Figures 1 and 2 about here.]

Moreover, perhaps the Inter-American Commission renders adverse decisions and the Court renders adverse judgments in states that are more democratic, as there is a greater likelihood of compliance when democratic institutions are strong. Or, adverse Commission decisions and Court judgments may be rendered more often in states with more robust civil society, as there will be a greater likelihood of domestic mobilization, and thereby compliance, in such states. Finally, perhaps the Commission and Court temper their judgments to accommodate the interests of actors responsible for supplying resources to the court (e.g. Carrubba, Gabel and Hankla 2008). Perhaps the Commission and Court consider their budgets (whether explicitly or implicitly) before ruling. For my purposes, the Commission and Court receive their funding from the Organization of American States, which receives funding from member state contributions based on GDP. As a result, one might expect states with a larger GDP to receive fewer adverse Commission decisions or Court judgments. On the other hand, perhaps the Commission or Court render judgments in states with greater capacity, as such states will have a greater ability to implement adverse decisions and judgments.

Figures 3 and 4 show that adverse Commission decision and Court judgment recipients are not necessarily more democratic countries, countries with more robust civil society, or countries with

higher/lower capacity. The upper left panel of Figures 3 and 4 show the number of adverse Commission decisions and Court judgments across empowerment rights. Empowerment rights data come from the Cingranelli, Richards and Clay (2014) empowerment rights index, which ranges 0-14 and is an additive index constructed from measures of several democratically-oriented rights including foreign movement, domestic movement, freedom of speech, freedom of assembly and association, workers' rights, electoral self-determination, and freedom of religion. The descriptive statistics show that adverse Commission decisions and Court judgments are more often made in states with relatively worse respect for democratically-oriented rights. The upper right panel of Figures 3 and 4 show the number of adverse Commission decisions and Court judgments across judicial independence using the Coppedge et al. (2016) indicator of high court independence, which captures how often the high court makes decisions that merely reflect government wishes, where higher values represent a more independent court. The results show that adverse Commission decisions and Court judgments are not rendered more often in countries with higher levels of judicial independence, but rather in countries with slightly less independent judiciaries.

The bottom left panel of Figures 3 and 4 show the number of adverse Commission decisions and Court judgments across the strength of civil society participation, using the Coppedge et al. (2016) indicator of civil society participation. This variable captures whether civil society organizations are routinely consulted by policymakers and the size of involvement of people in civil society organizations, among several other indicators. Higher values indicate a greater likelihood of civil society participation. The results show that adverse Commission decisions are made more often in states with weaker civil society and the Court renders judgments against states with varying levels of civil society participation. Finally, the bottom right panel of Figures 3 and 4 show the number of adverse Commission decisions and Court judgments across logged GDP per capita. The results show that adverse Commission decisions are more often made in states with slightly lower GDP per capita, while Court judgments are rendered in states with varying levels of capacity. Taken together, these descriptive statistics suggest that the adverse Commission decisions and Court judgments are not made more often in states with a greater likelihood of compliance. This suggests that the data

9

are likely to bias the results toward the null as Commission decisions and Court judgments are often made in states where human rights policy changes are less likely.

[Figures 3 and 4 about here.]

### 1.8.1  Matching

In an effort to further address selection concerns, I conduct an additional robustness check in which I pre-process the data using matching techniques and then estimate the negative binomial regression models. Doing so allows me to simulate a randomized experiment conditional on the observed covariates (Rubin 1974; Guo and Fraser 2010). The results are largely the same as the main models presented in the manuscript.

To begin, I preprocess the data prior to estimating each of the three models in the main manuscript. More specifically, I generate a dichotomous adverse Commission decision variable, in which an adverse Commission decisions takes on a value of 1 if a country received an adverse Commission decision in a given year (treatment) and 0 otherwise (control). Then, I perform nearest neighbor propensity score matching with the observed covariates included in the main models, yielding a dataset with similar units across the treatment and control groups. Table 9 shows the extent to which the matching exercise created more balance in the datasets across the treatment and control groups. After pre-processing, I estimate the negative binomial regression models described in the main manuscript. Table 10 displays coefficient estimates for the adverse Commission decision and adverse Court judgment variables for each of the three models described in the main manuscript. The coefficient estimates remain largely the same. In the first model, the adverse Commission decision variable is insignificantly associated with scarring torture, while the adverse Court judgment variable is negatively and significantly associated with scarring torture. In the second model, the adverse Commission decision and adverse Court judgment variable are insignificantly associated with clean torture. Finally, in the third model, the adverse Commission decision variable is positively and significantly associated with the number of formal complaints filed. These results combined with the descriptive statistics above lend greater support to the hy-

10

pothesized relationships.

[Tables 9 and 10 about here.]

# References

Carrubba, Clifford J., Matthew Gabel and Charles Hankla. 2008. "Judicial Behavior under Political Constraints: Evidence from the European Court of Justice." *American Political Science Review* 435–452(4):109.

Chayes, Abram and Antonia Handler Chayes. 1993. "On Compliance." *International Organization* 47(2):175–205.

Cingranelli, David, David L. Richards and Chad Clay. 2014. "The CIRI Human Rights Dataset.". Available at: `http://www.humanrightsdata.com`.

Conrad, Courtenay R., Daniel W. Jr. Hill and Will H. Moore. 2018. "Torture and the Limits of Democratic Institutions." *Journal of Peace Research* 55(1):3–17.

Conrad, Courtenay R., Jacqueline H.R. DeMeritt, Will H. Moore and Ryan Welch. 2013. *National Human Rights Institutions (NHRI) Institutional Data Project: (Beta) Coding Rules*. Charlotte, Denton, and Tallahassee: National Human Rights Institutions Data Project. Data available at: `http://psci.unt.edu/~demeritt/nhri/The_National_Human_Rights_Institutions_(NHRI)_Data_Collection_Project.html`.

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

Coppedge, Michael, John Gerring, Staffan Linberg, Svend-Erik Skaaning, Jan Teorell, David ALtman, Michael Berhard, M. Steven FIsh, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kelly McMann, Pamela Saxton, Daniel Pemstein, Jeffrey Staton, Brigitte Zimmerman, Rachel Sigman, Frida Andersson, Valeriya Mechkova and Farhad Miri. 2016. *V-Dem Codebook v6*. V-Dem Project.

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.

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

Fariss, Christopher J. 2018. "The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance." *British Journal of Political Science* 48(1):239–271.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Guo, Shenyang and Mark W. Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications*. Los Angeles: Sage.

Hawkins, Darren and Wade Jacoby. 2010. "Partial Compliance: A Comparison of the European and Inter-American American Courts for Human Rights." *Journal of International Law and International Relations* 6(1):35–85.

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

Lupu, Yonatan. 2013. "Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements." *International Organization* 67(3):469–503.

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

Nils Gleditsch, Peter Wallensteen, Mikael Erikson Margareta Sollenberg and Haavard Strand. 2002. "Armed Conflict, 1945-99: A New Dataset." *Journal of Peace Research* 39(5):615–637.

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

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688–701.

Welch, Ryan M. 2017. "National Human Rights Institutions: Domestic Implementation of International Human Rights Law." *Journal of Human Rights* 16(1):96–116.

Table 1: Descriptive Statistics

|  | Mean | Standard Deviation | Range |
|---|---|---|---|
| Commission | 0.389 | 1.39 | 0 - 11 |
| Court | 0.168 | 0.51 | 0 - 4 |
| Formal Complaint | 2.78 | 4.52 | 0 - 28 |
| Judicial Independence | 0.764 | 1.02 | -2.70 - 2.35 |
| HRO | 1.27 | 1.38 | 0 - 4 |
| Civil War | 0.148 | 0.356 | 0 - 1 |
| GDP (Logged) | 7.91 | 0.807 | 5.64 - 9.66 |
| Unstated | 3.97 | 8.00 | 0 - 68 |
| Clean | 1.68 | 3.12 | 0 - 19 |
| Scarring | 4.33 | 6.70 | 0 - 48 |
| Population (millions) | 23.09 | 40.18 | 0.264 - 190.70 |
| Restricted Access | 0.029 | 0.195 | 0 - 1 |
| Speech | 1.45 | 0.521 | 0 - 2 |
| Embeddedness | 1.77 | 0.660 | 0.704 - 3.66 |
| Repression (t-1) | 0.25 | 1.02 | -2.02 - 2.44 |

### Table 2: Model Results from Main Manuscript

| | Scarring | Clean | Formal Complaints |
|---|---|---|---|
| Commission | 0.037 | -0.005 | 0.053* |
| | (0.032) | (0.043) | (0.031) |
| Court | -0.257** | -0.074 | -0.218* |
| | (0.102) | (0.166) | (0.112) |
| Judicial Independence | 0.100 | -0.195* | -0.321*** |
| | (0.073) | (0.108) | (0.089) |
| NHRI | 0.357* | -0.237 | 0.082 |
| | (0.203) | (0.279) | (0.188) |
| Embeddedness | 0.259 | -0.031 | 0.080 |
| | (0.166) | (0.242) | (0.158) |
| HRO | -0.029 | 0.148* | -0.038 |
| | (0.062) | (0.086) | (0.068) |
| Civil War | -0.189 | -0.216 | 0.345 |
| | (0.236) | (0.317) | (0.276) |
| GDP (logged) | -0.308* | 0.023 | 0.154 |
| | (0.163) | (0.213) | (0.152) |
| Population (logged) | 0.573*** | -0.011 | 0.274** |
| | (0.101) | (0.123) | (0.110) |
| Speech | -0.072 | 0.089 | 0.132 |
| | (0.130) | (0.168) | (0.134) |
| Restricted Access | -0.298* | 0.432** | -0.119 |
| | (0.182) | (0.176) | (0.245) |
| Unstated | 0.006 | 0.009 | -0.005 |
| | (0.012) | (0.019) | (0.020) |
| Clean | 0.148*** | | 0.052 |
| | (0.026) | | (0.044) |
| Scarring | | 0.093*** | 0.067*** |
| | | (0.020) | (0.016) |
| Repression (t-1) | -0.186 | 0.045 | 0.196 |
| | (0.142) | (0.161) | (0.150) |
| Dispersion Parameter | 0.290 | 0.311 | 0.281 |
| | (0.063) | (0.107) | (0.082) |
| N | 162 | 162 | 162 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$.
Models estimated with robust standard errors. Two-tailed significance tests reported.

Table 3: Aggregate Torture Model (Ordered Logit)

|  | Torture |
|---|---|
| Court | -0.108 |
|  | (0.533) |
| Commission | 0.077 |
|  | (0.176) |
| Judicial Independence | 0.109 |
|  | (0.341) |
| NHRI | 0.867 |
|  | (0.648) |
| Embeddedness | -0.004 |
|  | (0.438) |
| HRO | -0.750** |
|  | (0.291) |
| Civil War | -0.771 |
|  | (1.01) |
| GDP (logged) | -0.181 |
|  | (0.492) |
| Population (logged) | 0.073 |
|  | (0.480) |
| Speech | 0.883** |
|  | (0.423) |
| Restricted Access | 2.28 |
|  | (1.67) |
| Repression (t-1)*Time | -0.078 |
|  | (0.094) |
| Repression (t-1) | 3.14* |
|  | (1.81) |
| Time | -0.091* |
|  | (0.054) |
| $R^2$ | 0.295 |
| N | 195 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$. Models estimated with robust standard errors. Two-tailed significance tests reported.

Table 4: Random Effects Negative Binomial Regression Models

| | Scarring | Stealth | Formal Complaints |
|---|---|---|---|
| Court | -0.208* | -0.098 | -0.102 |
| | (0.124) | (0.193) | (0.133) |
| Commission | 0.054* | -0.030 | 0.073** |
| | (0.032) | (0.057) | (0.034) |
| Judicial Independence | 0.122* | -0.254** | -0.358*** |
| | (0.076) | (0.119) | (0.079) |
| NHRI | 0.404** | -0.803*** | -0.091 |
| | (0.206) | (0.269) | (0.202) |
| Embeddedness | 0.193 | 0.141 | 0.122 |
| | (0.173) | (0.261) | (0.180) |
| HRO | -0.017 | 0.106 | -0.031 |
| | (0.063) | (0.130) | (0.068) |
| Civil War | -0.299* | 0.452* | 0.357* |
| | (0.164) | (0.272) | (0.187) |
| GDP (logged) | -0.168 | -0.147 | 0.113 |
| | (0.151) | (0.250) | (0.165) |
| Population (logged) | 0.519*** | 0.130 | 0.301** |
| | (0.089) | (0.195) | (0.107) |
| Speech | -0.099 | 0.175 | 0.103 |
| | (0.125) | (0.179) | (0.141) |
| Restricted Access | -0.176 | 0.578** | -0.119 |
| | (0.185) | (0.230) | (0.214) |
| Unstated | 0.001 | -0.010 | -0.020** |
| | (0.006) | (0.010) | (0.008) |
| Clean | 0.118*** | | 0.034 |
| | (0.018) | | (0.024) |
| Scarring | | 0.068*** | 0.065*** |
| | | (0.013) | (0.011) |
| Repression (t-1) | -0.224* | -0.111 | 0.170 |
| | (0.127) | (0.234) | (0.152) |
| Log Likelihood | -371.90 | -244.27 | -323.04 |
| N | 162 | 162 | 162 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$. Two-tailed significance tests reported.

Table 5: Country Fixed Effects Negative Binomial Regression Models

| | Scarring | Clean | Formal Complaints |
|---|---|---|---|
| Commission | 0.052 | -0.023 | 0.030 |
| | (0.037) | (0.052) | (0.034) |
| Court | -0.106 | -0.011 | -0.173* |
| | (0.124) | (0.186) | (0.107) |
| Judicial Independence | 0.047 | -0.101 | -0.094 |
| | (0.094) | (0.141) | (0.126) |
| NHRI | 0.554 | -0.254 | 0.559** |
| | (0.279) | (0.356) | (0.257) |
| Embeddedness | -0.001 | 0.251 | 0.065 |
| | (0.261) | (0.340) | (0.266) |
| HRO | -0.580** | 0.382 | -0.582* |
| | (0.226) | (0.324) | (0.369) |
| Civil War | -1.17*** | 0.439 | -0.354 |
| | (0.332) | (0.573) | (0.557) |
| GDP (logged) | -0.853*** | -0.028 | -0.061 |
| | (0.230) | (0.515) | (0.271) |
| Population (logged) | -6.62*** | -0.653 | -5.00** |
| | (1.72) | (2.77) | (1.99) |
| Speech | -0.239** | 0.108 | -0.094 |
| | (0.115) | (0.187) | (0.158) |
| Restricted Access | -0.227 | 0.599*** | -0.147 |
| | (0.115) | (0.188) | (0.189) |
| Unstated | -0.006 | -0.017 | -0.012 |
| | (0.010) | (0.017) | (0.016) |
| Clean | 0.128*** | | 0.031 |
| | (0.025) | | (0.035) |
| Scarring | | 0.084*** | 0.056*** |
| | | (0.019) | (0.016) |
| Repression (t-1) | -0.106 | -0.779** | -0.354 |
| | (0.257) | (0.314) | (0.288) |
| Log Likelihood | -343.10 | -224.96 | -303.62 |
| N | 162 | 162 | 162 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$. Two-tailed significance tests reported.

Table 6: Year Fixed Effects Negative Binomial Regression Models

| | Scarring | Clean | Formal Complaints |
|---|---|---|---|
| Commission | 0.017 | 0.001 | 0.049* |
| | (0.032) | (0.053) | (0.033) |
| Court | -0.190* | -0.132 | -0.207* |
| | (0.109) | (0.172) | (0.107) |
| Judicial Independence | 0.188*** | -0.139 | -0.200** |
| | (0.069) | (0.101) | (0.091) |
| NHRI | 0.343* | -0.211 | 0.081 |
| | (0.198) | (0.249) | (0.206) |
| Embeddedness | 0.669*** | 0.157 | 0.353* |
| | (0.158) | (0.217) | (0.206) |
| HRO | -0.095* | 0.153** | -0.076 |
| | (0.063) | (0.075) | (0.071) |
| Civil War | -0.441** | -0.291 | -0.028 |
| | (0.198) | (0.273) | (0.233) |
| GDP (logged) | -0.608*** | -0.124 | -0.088 |
| | (0.136) | (0.195) | (0.170) |
| Population (logged) | 0.826*** | 0.135 | 0.532*** |
| | (0.099) | (0.152) | (0.130) |
| Speech | -0.250* | 0.142 | -0.058 |
| | (0.129) | (0.209) | (0.141) |
| Restricted Access | -0.091 | 0.536* | -0.105 |
| | (0.219) | (0.298) | (0.229) |
| Unstated | -0.002 | 0.013 | -0.115 |
| | (0.008) | (0.015) | (0.011) |
| Clean | 0.133*** | | 0.081** |
| | (0.026) | | (0.037) |
| Scarring | | 0.069*** | 0.037*** |
| | | (0.024) | (0.013) |
| Repression (t-1) | -0.085 | 0.060 | 0.326** |
| | (0.143) | (0.167) | (0.165) |
| Log Likelihood | -355.55 | -236.93 | -312.87 |
| N | 162 | 162 | 162 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$. Two-tailed significance tests reported.

Table 7: Results from Models with Standard Errors Clustered on Country

| | Scarring | Stealth | Formal Complaints |
|---|---|---|---|
| Court | -0.256*** | -0.074 | -0.218** |
| | (0.081) | (0.197) | (0.033) |
| Commission | 0.037 | -0.005 | 0.054* |
| | (0.046) | (0.040) | (0.033) |
| Judicial Independence | 0.100 | -0.195 | -0.321*** |
| | (0.086) | (0.182) | (0.114) |
| NHRI | 0.357* | -0.237 | 0.082 |
| | (0.215) | (0.344) | (0.193) |
| Embeddedness | 0.259 | -0.031 | 0.080 |
| | (0.166) | (0.273) | (0.175) |
| HRO | -0.029 | 0.148** | -0.038 |
| | (0.064) | (0.069) | (0.062) |
| Civil War | -0.189 | -0.216 | 0.345 |
| | (0.224) | (0.340) | (0.289) |
| GDP (logged) | -0.308* | 0.023 | 0.154 |
| | (0.173) | (0.223) | (0.134) |
| Population (logged) | 0.573*** | -0.011 | 0.274** |
| | (0.129) | (0.199) | (0.112) |
| Speech | -0.072 | 0.089 | 0.132 |
| | (0.110) | (0.157) | (0.141) |
| Restricted Access | -0.298 | 0.432** | -0.119 |
| | (0.204) | (0.194) | (0.151) |
| Unstated | 0.006 | 0.009 | -0.005 |
| | (0.015) | (0.016) | (0.020) |
| Clean | 0.148*** | | 0.052 |
| | (0.033) | | (0.039) |
| Scarring | | 0.093*** | 0.067** |
| | | (0.026) | (0.016) |
| Repression (t-1) | -0.186 | 0.045 | 0.196 |
| | (0.142) | (0.218) | (0.195) |
| Log Likelihood | -377.81 | -241.98 | -324.34 |
| N | 162 | 162 | 162 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$. Models estimated with clustered standard errors on country. Two-tailed significance tests reported.

Table 8: Effect of Commission Decisions on All Commission Decision Recipients

| | Scarring | Stealth | Formal Complaints |
|---|---|---|---|
| Court | -0.209** | -0.132 | -0.135 |
| | (0.108) | (0.167) | (0.114) |
| Commission | 0.060 | -0.020 | 0.043 |
| | (0.048) | (0.042) | (0.033) |
| Judicial Independence | -0.005 | -0.294*** | -0.327*** |
| | (0.066) | (0.089) | (0.078) |
| NHRI | 0.158 | -0.306* | -0.010 |
| | (0.143) | (0.167) | (0.156) |
| Embeddedness | 0.037 | 0.013 | 0.135 |
| | (0.117) | (0.134) | (0.156) |
| HRO | -0.001 | 0.023*** | 0.004 |
| | (0.008) | (0.008) | (0.009) |
| Civil War | -0.241 | -0.104 | 0.025 |
| | (0.200) | (0.246) | (0.237) |
| GDP (logged) | -0.086 | 0.108 | -0.025 |
| | (0.134) | (0.172) | (0.115) |
| Population (logged) | 0.302*** | 0.103 | 0.194** |
| | (0.086) | (0.104) | (0.085) |
| Speech | -0.060 | -0.054 | 0.174 |
| | (0.108) | (0.136) | (0.112) |
| Restricted Access | -0.422*** | 0.231 | 0.148 |
| | (0.096) | (0.133) | (0.194) |
| Unstated | 0.026** | 0.020 | 0.002 |
| | (0.012) | (0.018) | (0.018) |
| Clean | 0.059*** | | 0.009 |
| | (0.014) | | (0.016) |
| Scarring | | 0.076*** | 0.084*** |
| | | (0.014) | (0.015) |
| Repression (t-1) | -0.422*** | -0.062 | 0.018 |
| | (0.096) | (0.133) | (0.112) |
| Log Likelihood | -504.77 | -341.32 | -422.57 |
| N | 209 | 209 | 209 |

NOTES: Parameter estimates and standard error reported. Statistical significance: ***$p < .01$, **$p < .05$, *$p < .10$. Models estimated with robust standard errors. Two-tailed significance tests reported. Formal complaint model results are statistically significant using a one-tailed significance test.

Figure 1: Number of Adverse Commission Decisions across Physical Integrity Rights
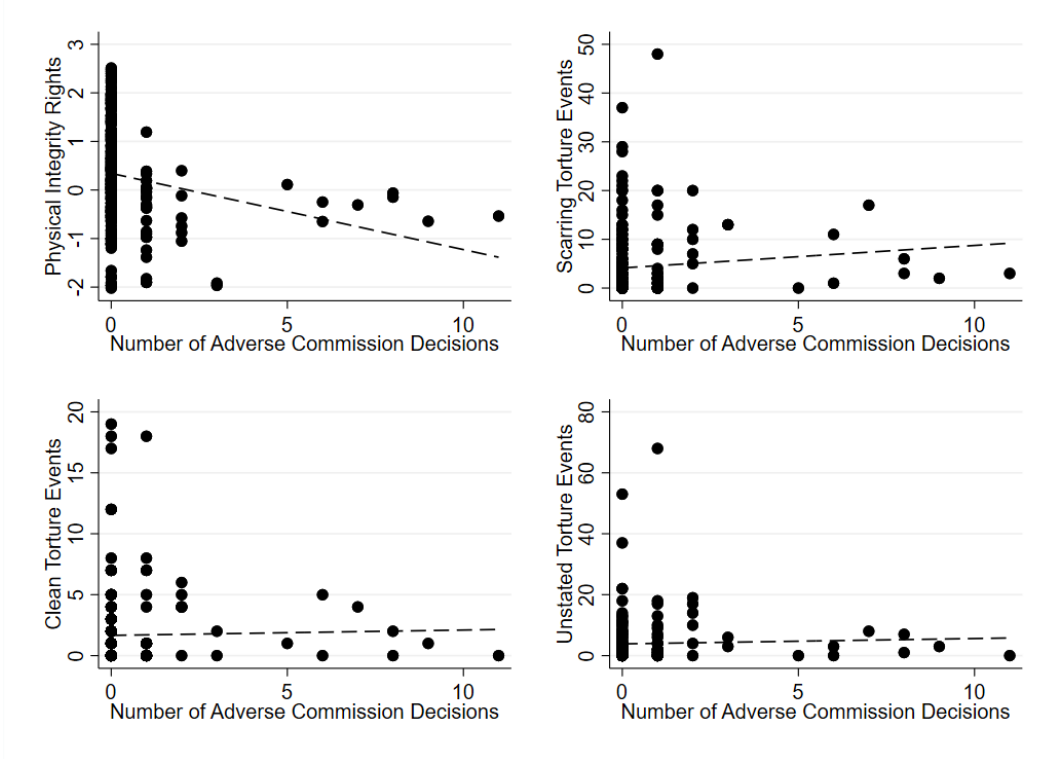


NOTES: Dashed lines display fitted values

Table 9: Propensity Score Balance

| | Before Matching | | After Matching | |
| --- | --- | --- | --- | --- |
| | Mean Treated | Mean Control | Mean Treated | Mean Control |
| Commission Decisions (Model 1) | 0.336 | 0.179 | 0.319 | 0.300 |
| n | 137 | 36 | 35 | 35 |
| Commission Decisions (Model 2) | 0.318 | 0.183 | 0.286 | 0.278 |
| n | 137 | 36 | 34 | 34 |
| Commission Decisions (Model 3) | 0.381 | 0.163 | 0.366 | 0.308 |
| n | 137 | 36 | 35 | 35 |

NOTES: Estimated average propensity score for adverse Commission decision variable for the treated and control groups for each model. Model 1 is the model predicting scarring torture, model 2 is the model predicting clean torture, and model 3 is the model predicting formal complaints. The estimated propensity scores are presented before and after data preprocessing with matching to show the extent to which balance occurs.

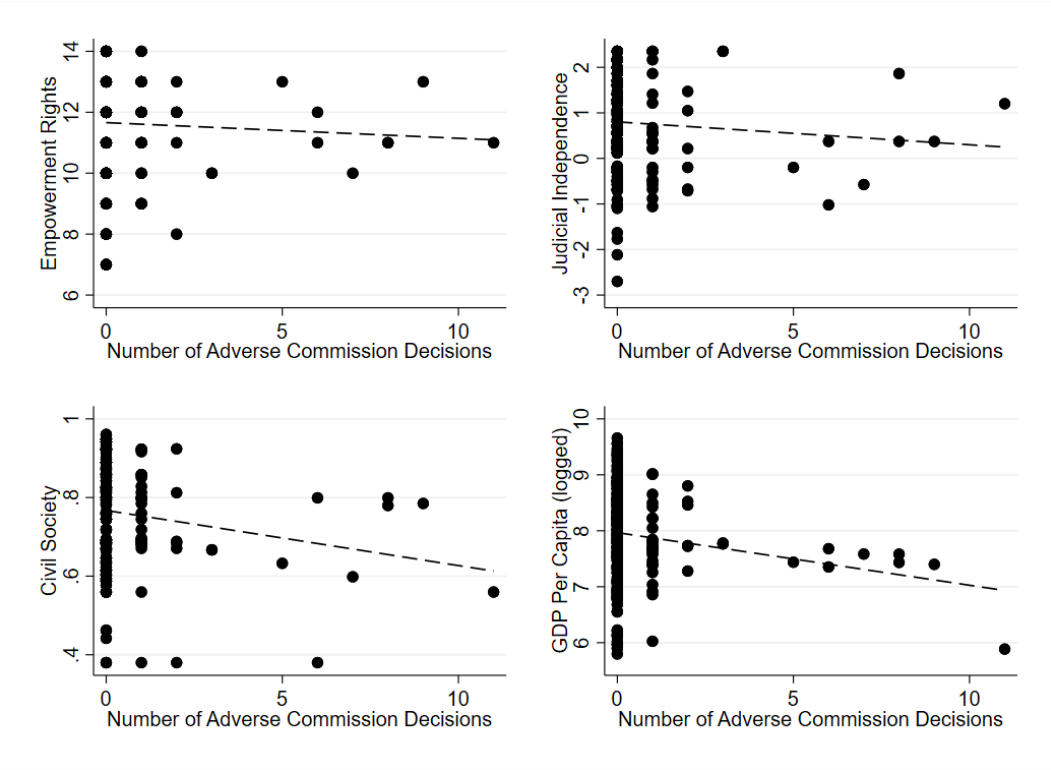Figure 2: Number of Adverse Court Judgments across Physical Integrity Rights
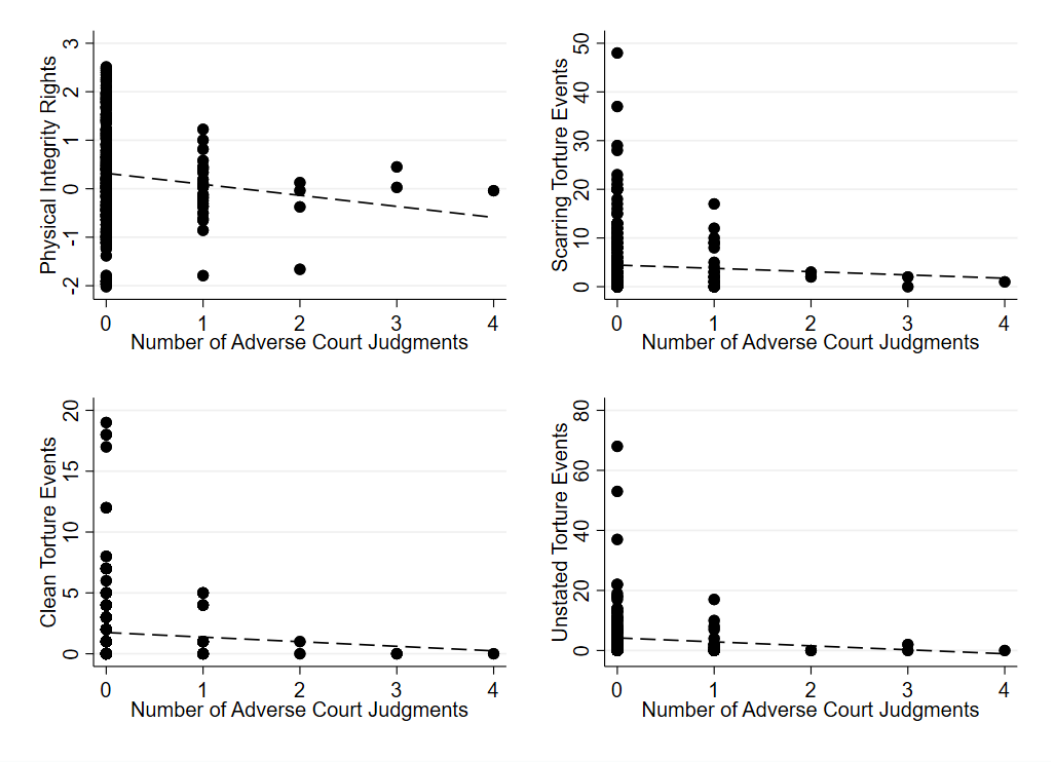


NOTES: Dashed lines display fitted values

Table 10: Coefficient Estimates for Models Using Treatment Matched Datasets

| Model 1 (Scarring torture) | Coefficient Estimate | P-Value |
|---|---|---|
| Commission Decisions | 0.038 | 0.227 |
| Court Judgments | -0.269 | 0.024 |
| Model 2 (Clean Torture) | Coefficient Estimate | P-Value |
| Commission Decisions | -0.003 | 0.947 |
| Court Judgments | 0.126 | 0.552 |
| Model 3 (Formal Complaints) | Coefficient Estimate | P-Value |
| Commission Decisions | 0.075 | 0.053 |
| Court Judgments | 0.003 | 0.982 |

NOTES: Coefficient estimates and their p-values for the primary variables of interest in the model (adverse Commission decisions and adverse Court judgments).
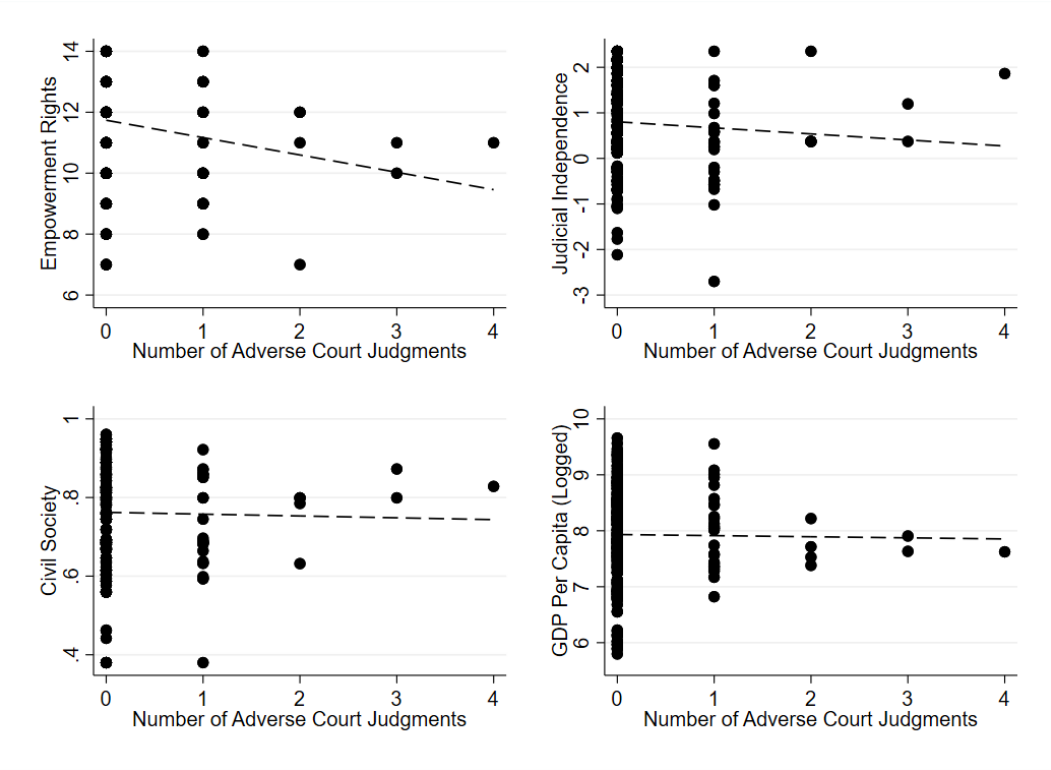
Figure 3: Number of Adverse Commission Decisions across Democratic Institutions, Civil Society, and GDP



NOTES: Dashed lines display fitted values

Figure 4: Number of Adverse Court Judgments across Democratic Institutions, Civil Society, and GDP



NOTES: Dashed lines display fitted values