Are New Democracies Better Human Rights Compliers?

Sharanbir Grewal and Erik Voeten

Abstract Recent scholarship finds that new democracies are more likely than established democracies to make binding commitments to international human rights institutions. Are new democracies also better at following through on these commitments? Stated differently, does their greater willingness to join international institutions reflect a genuine commitment to human rights reform or is it just "cheap talk?" We analyze this question using a new data set of more than 1,000 leading European Court of Human Rights (ECtHR) cases. Since new democracies face judgments that are more difficult to implement than established democracies, we employ a genetic matching algorithm to balance the data set. After controlling for bureaucratic and judicial capacity, we find that new democracies do implement similar ECtHR judgments initially more quickly than established democracies, but this effect reverses the longer a judgment remains pending. Although new democracies have incentives to implement judgments quickly, they sometimes lack checks and balances that help ensure implementation should an executive resist.

A growing literature asserts that new democracies are more likely than stable democracies to make binding commitments to international human rights institutions. Governments in new democracies may wish to "lock-in" reforms for future governments that could be less respectful of individual rights.[1] New democracies also have stronger incentives to credibly communicate to domestic and international audiences that they are sincere about human rights reform.[2] These governments are therefore more willing to pay the sovereignty costs associated with delegating authority to an international court with compulsory jurisdiction than are governments in established democracies.

Are new democracies also more likely to implement the judgments of international human rights courts? The theoretical arguments imply as much. If a government truly intends to lock in rights improvements, then it ought to implement international court

The authors would like to thank Karen Alter, Liesbeth Hooghe, Yonatan Lupu, Gary Marks, Michael Zurn, and two anonymous reviewers for helpful comments. An earlier version was presented at the 2013 Conference on European Studies (Amsterdam) and the 2013 annual Political Science Association meeting. This research was supported in part by the Lisa J. Raines Research Fellowship and the Carroll Fellows Initiative.

1. For example, Moravcsik 2000.

2. See, for example, Pevehouse 2002; Mansfield and Pevehouse 2006; Simmons and Danner 2010; Risse, Ropp, and Sikkink 1999, 8–9; Guzman 2008; and Elsig, Milewicz, and Stürchler 2011.

International Organization 69, Spring 2015, pp. 497–518
© The IO Foundation, 2015
doi:10.1017/S0020818314000435

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

judgments quickly. Moreover, ratification of a human rights treaty would not provide a credible signal if it were followed by blatant negligence in implementation. We should therefore expect quicker implementation among new democracies compared with mature ones. Whether this is so also has important normative implications for the international human rights regime, which is often criticized for mattering most where it is needed least: in stable democracies.[3]

We investigate this question using a new data set of more than 1,000 European Court of Human Rights (ECtHR) judgments. The ECtHR is a useful context given the large number of judgments and the mixture of new and established democracies among its forty-seven member states. Moreover, there is considerable variation in the implementation of ECtHR judgments.[4]

There are two important inferential obstacles to answering our question. First, new democracies suffer from infringements that are more difficult to remedy. We address this issue by matching the judgments on a large number of observable characteristics, such as the severity of the violations and the type of remedies required to comply with the judgment. This preprocessing creates a sample of comparable judgments across stable and new democracies.[5]

Second, implementing international legal obligations requires not just political incentives but also institutional capacity.[6] Once we control for differences in institutional capacity and account for varying case characteristics, we find that new democracies initially implement judgments more quickly than established democracies. However, this effect disappears the longer a judgment remains pending. Most cases that continue to be pending after a decade or more come from new democracies.

Our interpretation of this finding is that although new democracies have incentives to implement judgments quickly, they sometimes lack checks and balances that help ensure implementation should an executive resist. Checks and balances may slow down implementation, but where executives are hesitant independent courts or legislatures make implementation more likely. Thus new democracies are initially more responsive to ECtHR judgments but the impact of the court is more ensured in stable democracies with adequate checks and balances.

Our research design differs from most existing studies, which compare the human rights performance of states that have and have not ratified a treaty. We analyze whether there are systematic differences in how different states that have all ratified a treaty implement their treaty obligations. Such analyses will become increasingly important as many global treaty regimes approach near-universal membership and as human rights courts and semi-judicial bodies become increasingly prolific.

Examining questions such as ours within the traditional design would require the simultaneous evaluation of two treatment effects: ratification and regime type. We

3. See Hafner-Burton and Tsutsui 2007; and Neumayer 2005.
4. See Von Staden 2009; Hillebrecht 2010; and Hawkins and Jacoby 2010.
5. See Ho et al. 2007; and Diamond and Sekhon 2013.
6. See Chayes and Chayes 1993 and 1995.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

offer guidance on how to tackle the distinct causal inference problems that accompany our research design. Our primary concern is no longer that treaty ratifiers are systematically different from nonratifiers, but that compliance tasks vary for each ratifier. If good data about the nature of the compliance tasks are available, then matching offers a promising technique to compare the behavior of different states on similar implementation challenges, although matching does not control for unobservable differences. In the conclusion, we offer thoughts on other areas where such a research design could be profitably exploited.

Democratization and Compliance with International Court Judgments

Moravcsik argues that leaders of democratizing states have incentives to lock in human rights reform by committing their countries to binding international institutions, in particular the ECtHR. In new democracies, governments may fear that future governments will reverse human rights reform.[7] Such concerns may supersede worries about sovereignty costs. Governments in established democracies, by contrast, worry less about future governments undermining human rights. The logic behind Moravcsik's argument suggests that new democracies should also more quickly implement court judgments to actually entrench reforms, although this has not been subjected to empirical inquiries.

Commitments to international human rights courts may also signal to skeptical domestic and international audiences that a government is serious about respecting human rights.[8] Other democratic countries expect new democracies to commit to human rights[9] and donors, aid agencies, and trading partners are increasingly paying attention as well.[10] Moreover, investors may interpret a commitment to human rights as a signal that governments are willing to forego short-term benefits in exchange for long-term gains.[11] Yet domestic and international actors may be concerned about the credibility of human rights reform in new democracies.[12] Leaders may have time-inconsistent preferences: once coming to power, they face incentives to not implement announced reforms.[13] Weaker domestic accountability mechanisms may also fuel skepticism.[14] Given that human rights treaties are more effective when

7. Moravcsik 2000, 220. Landman (2005) finds evidence for the lock-in effect in his study of global human rights treaties, although Goodliffe and Hawkins (2006) find no support for the theory in their study of the Convention Against Torture (CAT).
8. See Pevehouse 2002; Mansfield and Pevehouse 2006; and Simmons and Danner 2010.
9. See Risse, Ropp, and Sikkink 1999, 8–9; Guzman 2008; and Hafner-Burton 2009.
10. See Richards, Gelleny, and Sacko 2001; Hafner-Burton 2009; Kelley 2004, 426; and Helfer and Voeten 2014.
11. Farber 2002.
12. Pevehouse 2002, 613.
13. Mansfield and Pevehouse 2006, 140.
14. Simmons and Danner 2010, 233.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

states accept enforcement mechanisms associated with those treaties,[15] states could increase the credibility of their commitment to human rights reform by accepting the ECtHR's compulsory jurisdiction.

Like lock-in, signaling provides stronger political incentives for implementation among new democracies. Established democracies may fail to implement ECtHR judgments without causing domestic or international audiences to doubt their ultimate commitment to human rights. New democracies do not have that luxury. Indeed, the (lack of) implementation of ECtHR judgments provides easily observable evidence that international actors such as the European Union and nongovernmental organizations (NGOs) use to evaluate new democracies.[16]

The empirical validity of this hypothesis is not straightforward to assess. First, judgments against countries with a more recent nondemocratic legacy may be more difficult to implement. As we shall show, judgments against new democracies are more likely to require legislative change, to concern multiple human rights violations, to group together multiple cases, and to involve more serious human rights violations such as torture. The severity of this caseload may well reflect the legacy of poorly functioning domestic legal institutions or inadequate control over police forces or other agencies that commit human rights abuses. Such cases will take longer to implement for any government. New democracies' more difficult caseloads may therefore conceal their greater incentives to comply.

Much of the compliance literature has focused on a different threat to causal inference: states that select into treaties may be systematically different from those that do not.[17] For example, Hill used matching to construct samples of similar countries that had and had not ratified a treaty.[18] Others use instrumental variables or structural equations (for example, Heckman selection models) to address the problem of selection into treaties.[19] Our question, however, is whether different states react differently to similar legal obligations. All states in our analysis have ratified the European Convention on Human Rights.[20] We hypothesize that the compliance pull of this legal commitment is greater for some states (new democracies) than others (established democracies).

Although the cases from new and established democracies differ systematically, there are cases that are sufficiently similar to allow for comparisons. For example, the cases of *Podkolzina v. Latvia*[21] and *Lykourezos v. Greece*[22] both concern violations of free elections (Article 3 of Protocol 1) for barring a candidate from holding a seat in parliament. In *Podkolzina,* Ingrīda Podkolzina was banned from running for

15. Cole 2012.
16. See Helfer and Voeten 2014.
17. See, for example, Von Stein 2005; Simmons and Hopkins 2005; Vreeland 2008; Hill 2010; and Lupu 2013.
18. Hill 2010.
19. See, for example, Simmons 2009; and Neumayer 2005.
20. Belarus is the only European state that has not ratified.
21. *Podkolzina v. Latvia*, ECHR decision of 46726/99 (2002) ECHR 2002-II.
22. *Lykourezos v. Greece*, ECHR decision of 33554/03 (2006) ECHR 2006-VIII.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

parliament on account of having insufficient knowledge of the Latvian language. Following the court's judgment, Latvia—a new democracy—amended its electoral law to delete this requirement just one month after the judgment, and completed its payment of just satisfaction to Podkolzina two months later. Even though language issues were controversial in Latvia, Latvian leaders had "an unflinching desire to join Western Europe."[23] By contrast, it took Greece more than two years to implement the judgment in *Lykourezos,* where the complainant had been forced to forfeit his parliamentary seat for having continued his law practice on the side.

Our empirical strategy is to first identify the characteristics that should make cases more difficult to implement. We then use a matching algorithm to construct a sample of cases for both new and old democracies that are similar on these characteristics.

There is a second problem that complicates inferences. Compliance is not just a function of political incentives. Managerialists have long argued that states with greater institutional capacity are better able to implement international legal obligations and therefore more likely to do so.[24] Most compliance theorists view managerial and political explanations as complementary.[25] A country's leadership may well decide that implementing an ECtHR judgment makes political sense, but many human rights violations concern structural problems with courts, prisons, the police, or other institutions. Such institutional reform may be quicker in countries with greater bureaucratic and judicial capacities. While there is no necessary link between democracy and capacity, empirically the newer democracies in Europe have relatively weaker capacities. We therefore must control for variation in capacities in our empirical model.

Finally, political institutions could affect implementation through different mechanisms than signaling and lock-in. A particularly prominent mechanism highlighted in the literature is the presence of checks and balances. The signaling and lock-in hypotheses specify incentives for executives. Yet there are also other actors in most democracies that influence the implementation of court judgments. When executives already have the incentives to comply, additional domestic veto players may actually slow down compliance.[26] On the other hand, in cases where executives are reluctant to comply, strong legislatures and independent judiciaries could force executives to engage in "begrudging compliance"[27] or even implement the judgment themselves through legislation or court rulings.[28] Stable democracies are more likely to have developed checks and balances. This may mean that stable democracies are more likely

23. Schmid 2008, 15. See also Galbreath 2006; and Hansson 2002.
24. See Chayes and Chayes 1993 and 1995.
25. See, for example, Tallberg 2002, 610; Hovi and Aakre 2010, 430; Von Staden 2009, 54; and Beach 2005, 124–26.
26. See Börzel, Hofmann, and Panke 2012; and Haverland 2000.
27. Hillebrecht 2012, 285.
28. See Helfer and Voeten 2014; Hillebrecht 2010 and 2012; Lupu 2013; and Stone Sweet and Keller 2008.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

than new democracies to implement judgments eventually, given that fewer mechanisms are available in a new democracy to force a reluctant executive to comply.

A Data Set of ECtHR Judgments

The ECtHR allows citizens from forty-seven Council of Europe (CoE) member states to charge that their government has violated a right protected by the European Convention on Human Rights. The Court has issued more than 30,000 judgments including many politically controversial opinions on the rights of gays to serve in the military, voting rights for prisoners, extradition of terrorism suspects, privacy rights of celebrities, the independence and efficiency of trials, property rights, and abortion rights.

Unlike the Inter-American Court of Human Rights, the ECtHR generally does not list specific requirements for implementation other than the monetary compensation that states ought to pay victims within three months of the judgment. The CoE's Committee of Ministers supervises the implementation of ECtHR judgments and may suggest other individual measures, such as reopening judicial proceedings. Other judgments require more general measures, including legislative change and changes in domestic court jurisprudence. If all aspects of a judgment have been implemented to the satisfaction of the Committee of Ministers, it adopts a final resolution outlining the various individual and general measures that have been taken, thereby closing the case. If not satisfied, the committee may adopt an interim resolution detailing what a state would need to accomplish in order to satisfactorily implement a judgment.

Case Selection

An initial list of cases was created by searching for all ECtHR judgments between 1 January 1960 and 31 December 2006 in the CoE's online Human Rights Documentation (HUDOC) database.[29] We limit our study to "lead cases," which are the first to reveal "a new structural/general problem in a respondent state and which thus require the adoption of new general measures."[30] Lead cases are contrasted with follow cases, which result from the same structural issue. For instance, *Aksoy v. Turkey*[31] is a lead case with 273 follow cases, all of which concern violations by Turkish security forces. The CoE's Committee of Ministers adopts only one resolution for lead and follow cases so we cannot consider follow cases independent observations. For cases that were still pending, the CoE's State of Execution website[32]

29. Available at <http://cmiskp.echr.coe.int/tkp197/search.asp?skin=hudoc-en>, accessed 21 May 2014.
30. Council of Europe 2010, 29.
31. *Aksoy v. Turkey*, ECHR decision of 21987/93 (1996) ECHR 1996-VI.
32. Available at <http://www.coe.int/t/dghl/monitoring/execution/Reports/pendingCases_en.asp>, accessed 21 May 2014. There were 123 judgments of importance level 1 or 2 that had violations of the

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

provides data on which pending cases are lead cases and which are follow cases.[33] We further eliminated cases that are assigned importance level 3 by the ECtHR. These are judgments "that simply apply existing case-law, friendly settlements and strike outs (unless raising a particular point of interest)."[34] We did, however, include importance level 3 judgments that had follow cases of importance level 1 or 2.[35]

Dependent Variable

This leaves us with 1,056 cases. Of these, 846 (80.1 percent) had received final resolutions by 22 September 2012. Since justice delayed is justice denied, we examine time to implementation as our key dependent variable. More specifically, we coded the time between the date of the court's judgment and the date the Committee of Ministers adopted a formal resolution closing the case. This gives us a much richer measure of implementation than simply evaluating whether countries complied eventually.

Of the cases that were formally resolved, the average time until resolution was 1,563 days. This average hides tremendous variation. The shortest-lasting case was *Quaranta v. Switzerland*,[36] receiving a resolution within six months of the judgment, and the longest was *F.C.B. v. Italy*,[37] finally receiving a resolution after 7,322 days.

We also read each resolution and coded the date the country took its last measure to implement a judgment. This provides an alternative measure of duration to implementation. Unfortunately, this date could not be established for about 10 percent of the cases, which could introduce bias.[38] Moreover, the dates were not reported systematically, thus we cannot be sure that the last date noted in the resolution was indeed the last measure needed to satisfy the Committee of Ministers.[39] We therefore use this latter measure for robustness checks only.

convention but did not have any interim or final resolutions nor were found on the State of Execution website. We are thus uncertain of whether these cases are lead or follow, and these cases were excluded from statistical analysis.

33. There are three exceptions to these general coding rules. First, in cases in which the lead case happened to be concluded with a friendly settlement, such as in *Faulkner v. the United Kingdom,* we instead select the next nonfriendly settlement case in the final resolution. Similarly, for lead cases whose judgments were made by the European Commission and not the ECtHR, such as *Govell v. the United Kingdom,* we used the next ECtHR case. Finally, in cases with multiple respondent nations, a separate "case" was created for each country in which the court identified a structural problem.

34. HUDOC FAQ, 3. Available at <http://www.echr.coe.int/Documents/HUDOC_FAQ_ENG.pdf>, accessed 21 May 2014.

35. This suggests initial misclassification by the ECtHR. There were thirty-seven such cases. For example, see *A.A.U. v. France*, ECHR decision of 44451/98 (2001).

36. *Quaranta v. Switzerland*, ECHR decision of 12744/87 (1991) ECHR 1991-A205.

37. *F.C.B. v. Italy*, ECHR decision of 12151/86 (1991) ECHR 1991-A208-B.

38. For example, missing dates are much more likely for judgments that require only individual measures (35 percent) than for judgments that require general (more easily observable) measures (5 percent). The difference between both measures is not significantly correlated with whether a state is a new democracy.

39. For example, a country may adopt legislation but the committee may need to see the legislation implemented before it is satisfied. We would observe the date of the legislation only, not the date of implementation measures.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

Independent Variable

We measure the independent variable in two ways. First, we adopt Moravcsik's definition of a stable democracy as any country that has consistently had a Polity score[40] of 6 or higher for thirty consecutive years. New democracies are countries with Polity scores of 6 or higher but which have not yet enjoyed this status for thirty years. Since we have almost no cases against countries with Polity scores less than 6, we exclude these from the data.

Our second measure requires that stable democracies have had a Polity score of 6 for ten years and currently have a perfect Polity score of 10. The reasoning is that relatively recent democratizers with imperfect Polity scores may not have fully stabilized. By these measures, 46 percent and 40 percent of judgments, respectively, were against new democracies.[41] We have no clear preference for either measure but we wish to assess the robustness of our findings to what both seem reasonable definitions of new and stable democracies.

Matching on Case Characteristics

A first threat to causal inference is that the cases against new democracies are substantially different from cases against stable democracies. This is especially problematic if the kinds of cases more frequently found against new democracies are also characteristics that make implementation more difficult for any state. Figure 1 provides ample reason for concern. The figure presents the differences in the mean values for key characteristics of judgments against new and stable democracies.[42] For example, the X at .07 indicates that in the unmatched sample cases against new democracies were 7 percentage points more likely to involve Articles 2 or 3, which concern the right to life and inhumane treatment (torture).

First, on average, cases against new democracies require more extensive reforms. Judgments against new democracies more frequently required legislative change (8 percentage points), new jurisprudence by domestic courts, executive and/or administrative actions, and practical measures such as renovating prisons or hiring judges.

Second, cases against new democracies on average concern violations of more convention articles and have more follow cases than cases against stable democracies.

Third, the substance of the human rights violations committed by new democracies differs in potentially consequential ways. Twelve percent of cases against new democracies concerned violations of Articles 2 or 3. This is true for only 5 percent of cases against stable democracies. Such violations almost by definition strike at

40. Marshall, Gurr, and Jaggers 2013.

41. There are 144 judgments that are against new democracies by the first measure but not the second. There are eighty-six cases against countries that have had a minimal level of democracy for thirty years but had not yet reached full democracy.

42. Using the first definition.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

the heart of executive control over a society.[43] New democracies also have more violations of the right for an effective remedy before national authorities (Article 13) and property rights (Article 1 of Protocol 1). By contrast, new democracies have relatively fewer privacy (Article 8) and freedom of speech (Article 10) violations.

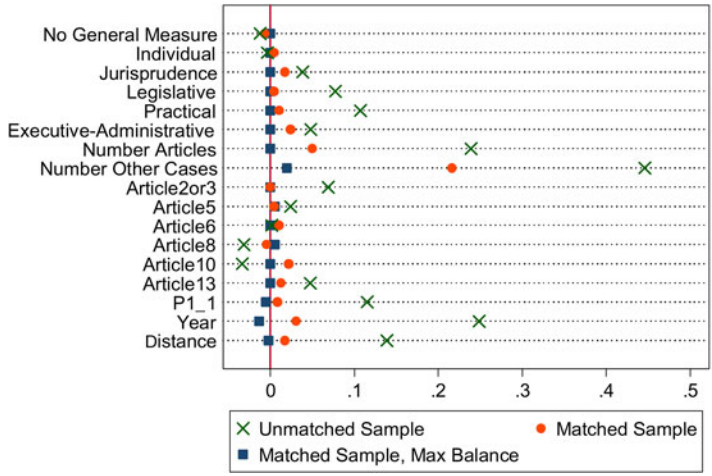


FIGURE 1. *Mean differences on case characteristics for new and stable democracies in unmatched and matched samples*

Finally, cases against new democracies on average were adopted about three years later than cases against stable democracies.[44] This is a consequence of new democracies joining the CoE on average at a later date. This may be consequential if monitoring or court practices have changed over time.

The most common strategy is to include these case characteristics as confounding variables in a multiple regression model. We show results from this conventional approach, but our preferred strategy is to preprocess the data using a matching algorithm.[45] Preprocessing makes the regression analysis less model-dependent by avoiding counterfactuals based on cases that are too dissimilar.[46] Given that we need to account for a large number of case characteristics (see Figure 1), this is a genuine concern. We apply a genetic matching algorithm,[47] which optimizes balance in the case characteristics between new and stable democracies according to predefined criteria. We can then apply conventional (weighted) regression models to the newly constructed sample to test our hypothesis.

43. Lupu and Voeten 2012, 421.
44. We transformed the year scale such that .1 reflects a difference of one year.
45. Ho et al. 2007.
46. Ibid.
47. See Sekhon 2011; and Diamond and Sekhon 2013.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

We follow two matching protocols.[48] In the first, most inclusive, the algorithm eliminates only control cases from stable democracies that do not have similar companion cases in the group of new democracies. Control cases from stable democracies that most resemble those from new democracies are more heavily weighted. In the second, more exclusive protocol, the algorithm drops both control and treatment cases that are outside the "convex hull."[49] This latter procedure minimizes the model dependence of follow-up analyses but also leaves fewer cases.

Figure 1 shows that the resulting samples are indeed more balanced: for all variables the mean differences between cases for new and stable democracies are closer to zero in the matched samples than in the unmatched sample. After the second procedure, which maximizes balance but throws out more cases, there are almost no differences between the observable characteristics of judgments from new and stable democracies (that is, all the dots are near zero).

The resulting data set may still be unbalanced when we consider interactions of the variables. For example, new democracies may have more Article 3 cases that require legislative change and have many follow cases. We therefore also examine a multidimensional measure of imbalance: L.[50] L drops from .71 in the unmatched sample to .43 in the matched sample, suggesting that the matching algorithm indeed reduced imbalance between the judgments issued against new and stable democracies. Similar results obtain for the alternative measure of stable democracy. Thus, matching has indeed increased the similarity in the observable characteristics of the judgments against new and stable democracies.

Are Democratizing States More Likely to Implement Quickly?

Our dependent variable is duration in days until implementation. Because some cases are still pending, the data are censored, necessitating a survival analysis. Figure 2 plots the Kaplan-Meier survival curves for the unmatched and matched samples.[51] The vertical axis gives the probability that a judgment will be implemented after a certain number of days (noted on the horizontal axis). In the unmatched data, new democracies take significantly longer than stable democracies to implement judgments. By contrast, when we consider similar judgments in the matched samples, the implementation speed between new and stable democracies becomes indistinguishable.

In the unmatched sample, stable democracies had implemented half of their judgments in about 1,500 days. This threshold is reached after about 2,000 days in the matched sample. This indicates that the matching procedure eliminated "easy" cases that were implemented quickly by stable democracies but that had no equivalent in new democracies. (Again, we do not match on duration, just on case characteristics.)

48. In both protocols, the population size was set at 5,000. R code is available from the authors.
49. See King and Zeng 2006 and 2007.
50. Iacus, King, and Porro 2011.
51. The plots are based on the first matching procedure.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

For example, freedom of speech cases constitute 10 percent of all cases against stable democracies in the unmatched data. This is only 4 percent in the matched data, given that new democracies attracted fewer such judgments. Such cases frequently require only individual measures, such as paying just satisfaction or reopening judicial proceedings. Indeed the percentage of judgments against stable democracies that require *only* individual measures drops from 18 percent to 14 percent in the matched sample.

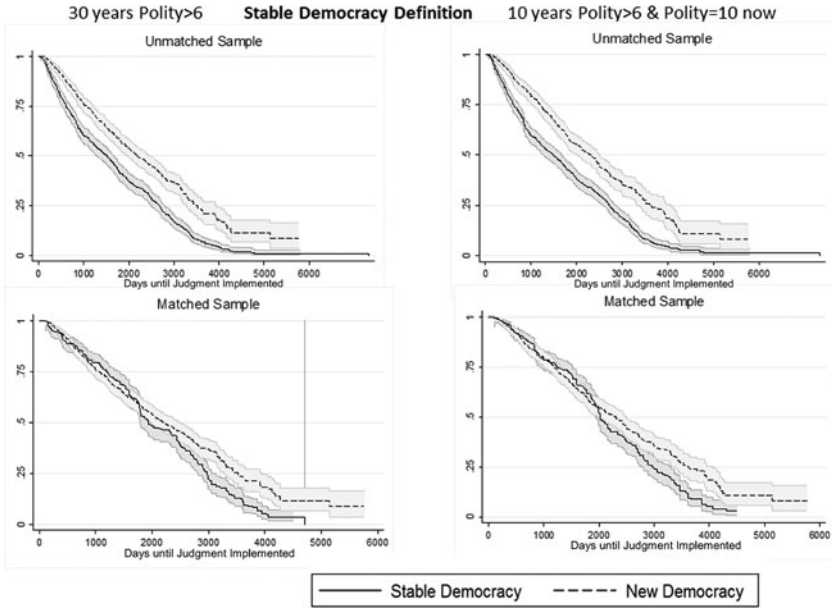


FIGURE 2. *kaplan-Meier Survival Functions in Matched and non-Matched Samples*

The graph also suggests that cases against stable democracies are less likely to remain pending indefinitely: judgments against new democracies that have already been pending for more than 2,000 days are less likely to be resolved than similarly long-lasting judgments against stable democracies.

Controlling for Variation in Bureaucratic Capacities

Figure 2 does not address the second threat to causal inference: that stable democracies may also have greater capacities to implement judgments quickly. The main managerial characteristic is the capacity of a country's bureaucracy and judiciary. Bureaucratic capacity is measured by the International Country Risk Guide (ICRG)'s bureaucratic quality variable.[52] This variable ranges from 0 to 4 and

52. Howell 2011.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

accounts for a bureaucracy's autonomy from political pressure and whether it has an established mechanism for recruitment and training. Judicial capacity is estimated by ICRG's LAW & ORDER variable.[53] The LAW & ORDER variable ranges from 0 to 6 and is itself a composite of LAW (the strength and impartiality of the legal system, 0 to 3) and ORDER (popular observance of the law, 0 to 3). Since the two variables are highly correlated, we combine them into a composite variable INSTITUTIONAL CAPACITY, which we rescale to the 0 to 1 range. The results are qualitatively identical if we take either variable separately.

Matching preprocesses the data but they should still be analyzed using (weighted) multiple regression analyses afterwards for proper inferences.[54] We estimate a semiparametric Cox hazard model. The hazard rate is as follows:

$$h_{ij}(t|\mathbf{X}_{ij}) = h_0(t)e^{\mathbf{X}_{ij}\boldsymbol{\beta}}$$

where $h_o(t)$is the baseline hazard function and X_{ij} are the values of the covariates for judgment j on country i. We estimate the model both on matched and unmatched samples, where the matched analyses use weights from the matching process.

A stringent assumption underlying Cox regression models is that the ratio of hazards is constant over time across individual judgments.[55] Figure 2 suggests that the effects of new democracy vary with time. The same is likely true for capacity. Capacity helps governments implement cases quickly initially but it should have little effect when governments resist implementation. Lack of capacity is not the reason that some Romanian or Turkish cases have been pending for more than a decade. We therefore interact both variables with the natural log of the duration of cases.

The models also include the case characteristics from Figure 1, although the results on the matched samples are robust to their exclusion. Throughout, we employ tests based on Schoenfeld residuals to assess whether the proportional hazard assumption is met. If not, we include interactions with time. Finally, we estimate stratified (by country) Cox regressions in order to account for unobserved country-specific fixed factors that may influence implementation. The results are robust to using shared frailty models to capture this heterogeneity instead.

Table 1 reports estimated hazard ratios, where values larger than 1 indicate that a higher value on the independent variable increases the probability of implementation and a value between 0 and 1 that higher values on the independent variable decrease the probability of implementation.

The hazard ratios for new democracy are significant and positive, whereas the interaction with time is significant and negative in all models but Model 2, which is the estimation on the unmatched sample where new democracy is defined using the "ten year" definition.

53. Ibid.
54. Ho et al. 2007.
55. See, for example, Box-Steffensmeier and Zorn 2001.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

TABLE 1. *Cox regression on duration to implementation for two measures of new democracy*

	Unmatched sample		*Matched sample*		*Matched sample, maximum balance*			
Stable democracy definition	*(1) 30 years*	*(2) 10 years*	*(3) 30 years*	*(4) 10 years*	*(5) 30 years*	*(6) 10 years*	*(7) 30 years*	*(8) 10 years*
NEW DEMOCRACY	10.10***	1.67	102***	130.2***	45.2**	54.3***	21.1***	59.5***
	(7.62)	(1.27)	(102)	(138)	(17.57)	(63.0)	(27.5)	(68.2)
NEW DEMOCRACY*TIME	0.75***	0.92	0.55***	0.51***	0.722**	0.57***	0.66***	0.57***
	(0.081)	(0.099)	(0.08)	(0.07)	(0.115)	(0.09)	(0.12)	(0.09)
INSTITUTIONAL CAPACITY	7.09e + 06***	210,287***	3.51e + 06***	169.6***	8.7e + 10***	451.7***	2.0e + 10***	1.4e + 06***
	(1.79e + 07)	(496,910)	(7.36e + 06)	(525,600)	(3.66e + 10)	(1.63e + 07)	(7.9e + 10)	(4.6e + 05)
INSTITUTIONAL CAPACITY*TIME	0.14***	0.21***	0.17***	0.23***	0.04***	0.19***	.055***	0.17***
	(0.05)	(0.07)	(0.07)	(0.09)	(0.02)	(0.09)	(0.03)	(0.08)
Case characteristics	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Observations	982	982	751	695	472	360	472	360

Notes: Entries are hazard ratios. Standard errors are in parentheses. *** $p < .01$, ** $p < .05$.

The substantive interpretation of the effects is complicated by the interactions with time. The constituent term is the effect of new democracy the day after the judgment, which is not our coefficient of interest. Figure 3 plots the relative hazard of a new democracy implementing a judgment at any given time.[56] The plots are based on

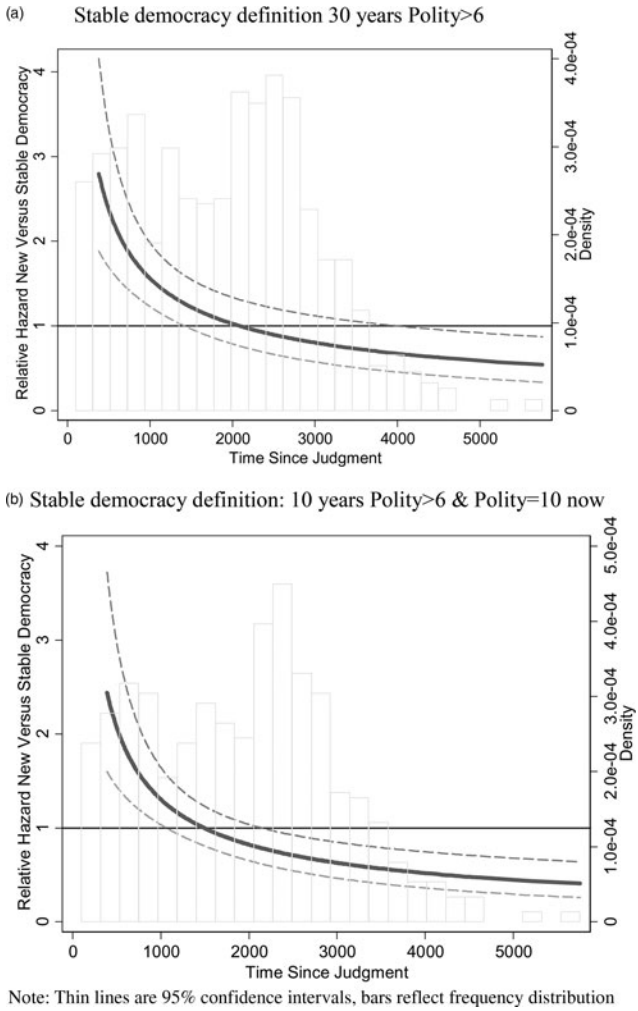


FIGURE 3. *Relative hazards that new versus stable democracies implement ECtHR judgments by time since judgment for different definitions of stable democracy*

56. We use the code supplied by Licht 2011.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

Models 3 and 4 (plots for the other models are similar). The relative hazard gives the change in the hazard of implementation between new and stable democracies. If the relative hazard equals 2, then new democracies are twice as likely as stable democracies to implement a judgment. The reverse is true if the relative hazard equals .5. The figure also plots the frequency distributions of cases in the respective matched samples.

The relative hazards for the first year are omitted because they involve relatively few cases (7 percent) and the hazards are very large. Between the first and the second year, a new democracy is two to three times more likely to implement a judgment than a stable democracy. The relative hazard equals 1 (no difference) after about four years in each model. If a judgment takes longer than four years to implement (about 40 percent of all judgments in the matched sample), stable democracies become more likely to implement. This is especially noticeable in the estimates from Model 4, which are based on a definition of a stable democracy as countries that had a Polity score of 10 but had only been a democracy for ten years at the time of judgment. After a judgment has remained pending for about a decade, stable democracies become twice as likely as new democracies to implement it. This concerns only a small proportion of all judgments (about 5 percent) but these are often the most controversial ones.

Figure 4a offers a similar plot for institutional capacity (based on Model 3). Since relative hazards have a less intuitive interpretation for continuous variables, we plot the combined coefficients.[57] The substantive interpretation of this coefficient is somewhat less straightforward and requires more computation given that a one-unit increase in institutional capacity goes beyond the data (it varies between .371 and 1 with a standard deviation of .19). A country that is one standard deviation above the mean is about twice as likely to implement an ECtHR judgment in the first year as a country with average institutional capacity.[58] Yet once a judgment has remained pending for more than five years, countries with greater capacity are no longer quicker in implementing judgments. The effect never goes below 0 in the observed range of cases, so institutional capacity does not slow down implementation but it simply becomes irrelevant. Cases are pending longer in low-capacity countries but there are probably other reasons that judgments have not been implemented after five or more years.

We suggested that this effect may be because of differences in checks and balances. This assertion is difficult to test directly given that the operationalization of new democracy through the Polity measure includes constraints on the executive. We assessed the plausibility of this mechanism using Henisz's measure of political

57. Ibid.

58. Institutional capacity is not the treatment variable in this study (that is, we did not balance the data with regard to capacity) and thus these estimates should not be given a causal interpretation.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

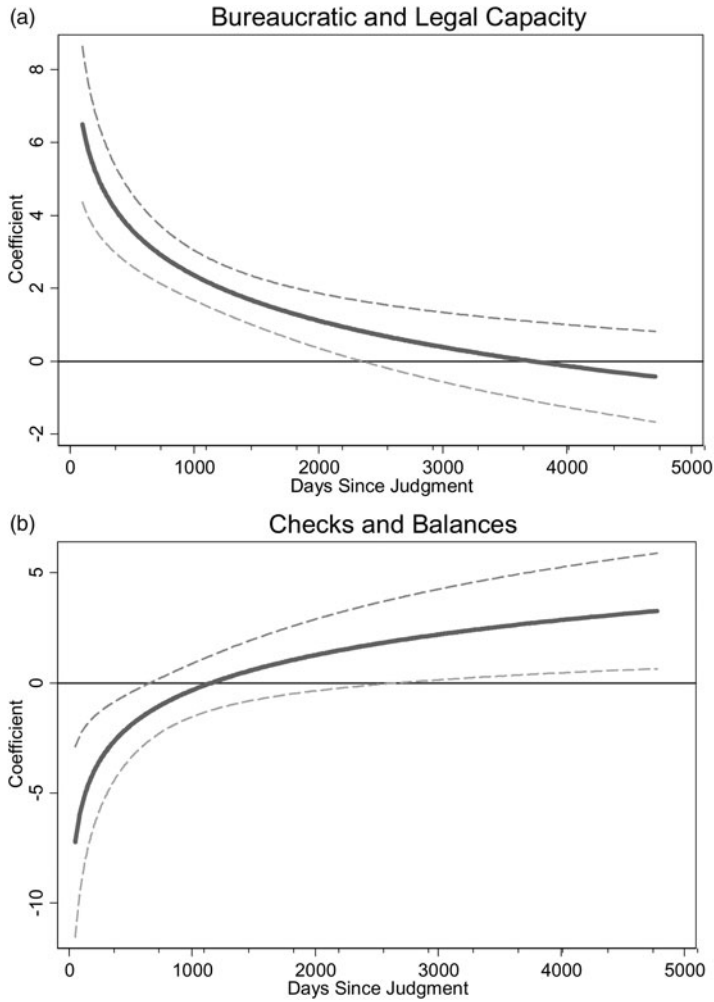


FIGURE 4. *Combined coefficients for bureaucratic and legal capacity (in figure a) and checks and balances (figure b). Thin lines are 95% confidence intervals.*[59]

constraints POLCONIII.[60] This measure combines information on the number of independent branches of government with veto power and the distribution of preferences within those branches. Thus, countries with many institutions that can exercise checks and balances and where those institutions are controlled by actors from different political parties receive higher scores.

59. Note that inside labels indicate years
60. Henisz 2000.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

Since we cannot match[61] on political constraints, we examine this using the unmatched data. We first reestimate Model 1 including POLCONIII in place of the new democracy dummy. Figure 4b offers the most important finding. Among judgments that are implemented quickly, fewer are from countries with high checks and balances. However, once a judgment remains pending longer, more checks and balances are correlated with speedier implementation. The effect of capacity is virtually indistinguishable from that in Figure 4a.

We replicated this analysis for all models estimated in Table 1 and found that checks and balances follow the same pattern in each model, consistent with our hypothesized interactive effect of checks and balances with time. The new democracy dummy is still significant and similarly signed (though smaller). This suggests that there is more to the new democracy variable's effect on implementation than just checks and balances. Of course, these findings are no more than suggestive about the mechanism through which new democracy has an interactive effect with time given that the relationship between new democracy and POLCONIII is likely endogenous.[62]

Robustness Checks

We replicated the analyses from Table 1 with our alternative dependent variable.[63] The results are similar: new democracies are significantly more likely to implement judgments quickly but this effect diminishes as judgments remain pending longer. This holds regardless of whether we imputed the missing values on the second dependent variable using information from our core dependent variable.

We examined several other judgment characteristics as control variables and in the matching algorithm: whether the case invited a separate opinion, whether it was referred to the Grand Chamber, and the court assigned importance of the case. These factors plausibly indicate legal controversies surrounding the ECtHR judgments. Including these factors in the matching algorithm or as control variables did not change our core findings; neither did eliminating the number of follow cases[64] from the set of matching characteristics.

We also examined several country-specific variables that could be confounding. First, we examined whether the new democracy effect was because of EU conditionality. However, a dummy for whether a country had applied for but not yet received EU membership was insignificant and did not alter our conclusions.

Second, gross domestic product (GDP) could help us assess whether stable democracies are less likely to implement quickly because they are relatively more

61. Unless we transform a continuous measure into a categorical one.

62. New democracies on average have about a standard deviation lower score than stable democracies on the POLCON measure.

63. All robustness checks are available from the authors.

64. This is a posttreatment factor so it may be problematic to condition on it. Still, the inclusion or exclusion leaves results unchanged.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

economically powerful and thus are less susceptible to reputational and material incentives.[65] We found no evidence for this proposition and inclusion of this variable did not alter our findings.

Third, perhaps new democracies disproportionally had left-wing governments that were quicker to implement judgments. We included an indicator for whether a country had a left-wing government (according to the Database of Political Institutions)[66] but found no evidence for this proposition.

Fourth, we restricted the analyses to those judgments adopted by the post Protocol XI court (since 1999) and found the same patterns. We also restricted the analysis to the post–Cold War analysis. This also did not affect the main results.

Finally, we included measures for the domestic human rights records of states: a state's physical integrity rights and an empowerment index (both from the Cingranelli-Richards (CIRI) Human Rights Dataset)[67]. Neither of these variables significantly affected implementation and the introduction of these variables (either alone or together) did not affect our estimates of the new democracy effect.

In short, the effect of new democracy on the implementation of ECtHR judgments is robust to alternative specifications and matching protocols. This is not a guarantee that a causal effect exists. Matching and multiple regression techniques control for only observables. There may be some unobservable process that makes cases against new democracies easier to implement than our estimates imply.

Conclusion

Controlling for institutional capacity, new democracies initially implement similar ECtHR judgments more quickly than stable democracies. This empirical finding is consistent with the theoretical insight that new democracies have greater incentives not only to bind themselves to international human rights courts but also to comply with their judgments because doing so enables them to signal their commitment to and/or lock in human rights reform. This counters the common criticism that the international human rights regime has its greatest effect in stable democracies with strong civil societies.

There is an important caveat to this finding: as a judgment remains pending longer, stable democracies become more likely to implement than new democracies. Our interpretation of this finding is that the political incentives to signal and lock in human rights reform may simply not be strong enough in some cases that are sensitive to executives. Stable democracies have fewer political incentives to act quickly but they do have checks and balances that ultimately (usually) ensure compliance. New democracies often lack such checks and balances. The UK resists many ECtHR judgments but it implements them, eventually. The same cannot always be said of Turkey and

65. See Börzel, Hofmann, and Panke 2012, 458; and Mbaye 2001, 263.
66. Beck et al. 2001.
67. Cingranelli, Richards, and Clay 2012.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

Romania, which have judgments that have remained pending for over a decade. Ultimately, then, there are limitations to the role that international human rights courts can play in democratizing states.

These findings raise further research questions regarding the relationship between checks and balances and democratization in the context of implementing international legal obligations. Our findings suggest that new democracies do have incentives to make costly signals but that the checks and balances that come with stable democracy matter to eventually force executives to comply begrudgingly. Nevertheless, in this research note we can ascertain only the plausibility of this complex set of relationships between domestic political institutions and compliance.

We also offer a contribution in research design. Most of the literature evaluates institutional effectiveness by comparing members and nonmembers. Yet much can be learned by evaluating how members of an institution respond differently to institutional demands. This design may become increasingly prominent as ratification of human rights agreements becomes near universal. Moreover, it is especially promising to test theories that institutions affect different states differently. Evaluating such claims in the traditional research design is complicated because there are two "treatments:" whether a state ratified a treaty and whether a state exhibits the conceptual feature hypothesized to enhance or reduce the institutional effect (in our case, being a new democracy).

A major challenge in our design is that not all institutional decisions are equally difficult to implement and that this variation may be correlated with our independent variable. We tackle this challenge by collecting information on a large number of judgment characteristics. We then match on these observable characteristics to create a sample of comparable cases. This makes our inferences more robust by ensuring that we are comparing the behavior of different countries on similar implementation tasks.

This design can potentially be applied to any institutional setting with multiple decisions or judgments. This includes other court-like settings, such as the Inter-American Court of Human Rights, the World Trade Organization's Dispute Settlement Understanding, the European Court of Justice, or investment tribunals. However, it may also be relevant in other settings where institutions make varying demands of states. For example, Piccone has gathered data on thousands of communications sent by UN human rights experts and the responses of states to them.[68] He also gathered information about the nature of the alleged human rights abuses. The design employed in this present study could plausibly be used to examine whether democracies indeed face greater costs of ignoring shaming.

A limitation is that while matching helps account for observable differences in implementation tasks, there may still be consequential unobserved differences between judgments issued against different countries. This issue can be tackled only if plausible instrumental variables or natural experiments can be found, which are rare in the context of international institutions. This limitation applies to all research designs.

68. Piccone 2012.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

References

Beach, Derek. 2005. Why Governments Comply: An Integrative Compliance Model That Bridges the Gap Between Instrumental and Normative Models of Compliance. *Journal of European Public Policy* 12 (1): 113–42.

Beck, Thorsten, George Clarke, Alberto Groff, Philip Keefer, and Patrick Walsh. 2001. New Tools in Comparative Political Economy: The Database of Political Institutions. *World Bank Economic Review* 15 (1):165–76.

Börzel, Tanja A., Tobias Hofmann, and Diana Panke. 2012. Caving in or Sitting It Out? *Longitudinal* Patterns of Noncompliance in the European Union. *Journal of European Public Policy* 19 (4):454–71.

Box-Steffensmeier, Janet M., and Christopher J.W. Zorn. 2001. Duration Models and Proportional Hazards in Political Science. *American Journal of Political Science* 45 (4):972–88.

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

———. 1995. *The New Sovereignty: Compliance with International Regulatory Agreements*. Cambridge, MA: Harvard University Press.

Cingranelli, David L., David L. Richards, and K. Chad Clay. 2012. The Cingranelli-Richards (CIRI) Human Rights Dataset. Available at <http://humanrightsdata.blogspot.com/p/data-documentation.html>. Accessed 19 November 2013.

Cole, Wade M. 2012. Human Rights as Myth and Ceremony? Reevaluating the Effectiveness of Human Rights Treaties, 1981–2007. *American Journal of Sociology* 117 (4):1131–71.

Council of Europe. 2010. Committee of Ministers. *Supervision of the Execution of Judgments of the European Court of Human Rights.* 4th Annual Report. Strasbourg, France: Council of Europe, Directorate General of Human Rights and Legal Affairs.

Diamond, Alexis, and Jasjeet S. Sekhon. 2013. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics* 95 (3):932–45.

Elsig, Manfred, Karolina Milewicz, and Nikolas Stürchler. 2011. Who Is in Love with Multilateralism? Treaty Commitment in the Post–Cold War Era. *European Union Politics* 12 (4):529–50.

Farber, Daniel A. 2002. Rights as Signals. *Journal of Legal Studies* 31 (1):83–98.

Galbreath, David J. 2006. European Integration Through Democratic Conditionality: Latvia in the Context of Minority Rights. *Journal of Contemporary European Studies* 14 (1):69–87.

Goodliffe, Jay, and Darren G. Hawkins. 2006. Explaining Commitment: States and the Convention Against Torture. *Journal of Politics* 68 (2):358–71.

Guzman, Andrew T. 2008. *How International Law Works: A Rational Choice Theory*. New York: Oxford University Press.

Hafner-Burton, Emilie M. 2009. *Forced To Be Good: Why Trade Agreements Boost Human Rights*. Ithaca, NY: Cornell University Press.

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):407–25.

Hansson, Ulf. 2002. The Latvian Language Legislation and the Involvement of the OSCE-HCNM: The Developments 2000–2002. *Global Review of Ethnopolitics* 2 (1):17–28.

Haverland, Markus. 2000. National Adaptation to European Integration: The Importance of Institutional Veto Points. *Journal of Public Policy* 20 (1):83–103.

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

Helfer, Laurence R., and Erik Voeten. 2014. International Courts as Agents of Legal Change: Evidence from LGBT Rights in Europe. *International Organization* 68 (1):77–110.

Henisz, Witold J. 2000. The Institutional Environment for Economic Growth. *Economics and Politics* 12 (1):1–31.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

Hill, Daniel W., Jr. 2010. Estimating the Effects of Human Rights Treaties on State Behavior. *Journal of Politics* 72 (4):1161–74.

Hillebrecht, Courtney. 2010. From Paper Tigers to Engines of Change: The Effect of Regional Human Rights Courts on Domestic Practice and Politics. PhD diss., University of Wisconsin-Madison.

———. 2012. Implementing International Human Rights Law at Home: Domestic Politics and the European Court of Human Rights. *Human Rights Review* 13 (3):279–301.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 15 (3):199–236.

Hovi, Jon, and Stine Aakre. 2010. Emission Trading: Participation Enforcement Determines the Need for Compliance Enforcement. *European Union Politics* 11 (3):427–45.

Howell, Llewellyn D. 2011. ICRG Methodology. Syracuse, NY: PRS Group. Available at <http://www.prsgroup.com/about-us/our-two-methodologies/icrg>. Accessed 19 November 2013.

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2011. Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association* 106 (493):345–61.

Kelley, Judith. 2004. International Actors on the Domestic Scene: Membership Conditionality and Socialization by International Institutions. *International Organization* 58 (3):425–57.

King, Gary, and Langche Zeng. 2006. The Dangers of Extreme Counterfactuals. *Political Analysis* 14 (2): 131–59.

———. 2007. When Can History Be Our Guide? The Pitfalls of Counterfactual Inference. *International Studies Quarterly* 51 (1):183–210.

Landman, Todd. 2005. *Protecting Human Rights: A Comparative Study*. Washington, DC: Georgetown University Press.

Licht, Amanda A. 2011. Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis. *Political Analysis* 19 (2):227–43.

Lupu, Yonatan. 2013. Legislative Veto Players and the Effects of International Human Rights Agreements. Paper presented at the Conference on International Law and Politics, September, Emory University.

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):413–39.

Mansfield, Edward D., and Jon C. Pevehouse. 2006. Democratization and International Organizations. *International Organization* 60 (1):137–67.

Marshall, Monty G., Ted Robert Gurr and Keith Jaggers. 2013. Polity IV Project: Political Regime Characteristics and Transitions, 1800–2010. Dataset Users' Manual. Center for Systemic Peace. College Park: University of Maryland. Available at <http://www.systemicpeace.org/inscr/p4manualv2012.pdf>. Accessed 19 November 2013.

Mbaye, Heather A.D. 2001. Why National States Comply with Supranational Law: Explaining Implementation Infringements in the European Union, 1972–1993. *European Union Politics* 2 (3): 259–81.

Moravcsik, Andrew. 2000. The Origins of International Human Rights Regimes: Democratic Delegation in Postwar Europe. *International Organization* 54 (2):217–52.

Neumayer, Eric. 2005. Do International Human Rights Treaties Improve Respect for Human Rights? *Journal of Conflict Resolution* 49 (6):925–53.

Pevehouse, Jon C. 2002. With a Little Help from My Friends? Regional Organizations and the Consolidation of Democracy. *American Journal of Political Science* 46 (3):611–26.

Piccone, Ted. 2012. *Catalysts for Change: How the UN's Independent Experts Promote Human Rights*. Washington, DC: Brookings Institution Press.

Richards, David L., Ronald D. Gelleny, and David H. Sacko. 2001. Money with a Mean Streak? Foreign Economic Penetration and Government Respect for Human Rights in Developing Countries. *International Studies Quarterly* 45 (2):219–39.

Risse, Thomas, Stephen C. Ropp, and Kathryn Sikkink, eds. 1999. *The Power of Human Rights: International Norms and Domestic Change*. Cambridge, UK: Cambridge University Press.

Schmid, Carol. 2008. Ethnicity and Language Tensions in Latvia. *Language Policy* 7 (1):3–19.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press

Sekhon, Jasjeet S. 2011. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. *Journal of Statistical Software* 42 (7):1–52.
Simmons, Beth A. 2009. *Mobilizing for Human Rights*. New York: Cambridge University Press.
Simmons, Beth A., and Allison Danner. 2010. Credible Commitments and the International Criminal Court. *International Organization* 64 (2):225–56.
Simmons, Beth A., and Daniel J. Hopkins. 2005. The Constraining Power of International Treaties: Theory and Methods. *American Political Science Review* 99 (4):623–31.
Stone Sweet, Alec, and Helen Keller, eds. 2008. *A Europe of Rights: The Impact of the ECHR on National Legal Systems*. Oxford, UK: Oxford University Press.
Tallberg, Jonas. 2002. Paths to Compliance: Enforcement, Management, and the European Union. *International Organization* 56 (3):609–43.
Von Staden, Andreas. 2009. Shaping Human Rights Policy in Liberal Democracies: Assessing and Explaining Compliance with the Judgments of the European Court of Human Rights. PhD diss., Princeton University.
Von Stein, Jana. 2005. Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance. *American Political Science Review* 99 (4):611–22.
Vreeland, James. 2008. Political Institutions and Human Rights: Why Dictatorships Enter into the United Nations Convention Against Torture. *International Organization* 62 (1):65–101.

https://doi.org/10.1017/S0020818314000435 Published online by Cambridge University Press