

University of Chicago Law School

Chicago Unbound

Journal Articles

Faculty Scholarship

2018

Treaties and Human Rights: The Role of Long-Term Trends

Adam S. Chilton

Eric A. Posner

Follow this and additional works at: https://chicagounbound.uchicago.edu/journal_articles



Part of the [Law Commons](#)

Recommended Citation

Adam S. Chilton & Eric Posner, "Treaties and Human Rights: The Role of Long-Term Trends," 81 Law & Contemporary Problems 1 (2018).

This Article is brought to you for free and open access by the Faculty Scholarship at Chicago Unbound. It has been accepted for inclusion in Journal Articles by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

TREATIES AND HUMAN RIGHTS: THE ROLE OF LONG-TERM TRENDS

ADAM S. CHILTON & ERIC A. POSNER*

I

INTRODUCTION

In recent years, several scholars who study human rights treaties have offered empirical evidence that after countries ratify these treaties, human rights improve.¹ This work contrasts with earlier work that had largely failed to find a positive correlation between treaty ratification and human rights outcomes.² The newer work relies on different datasets, methods, and research designs.

In this paper, we argue that the positive correlations between human rights treaties and outcomes found by this newer work may be due to failure to account for time trends. Human rights outcomes, as typically defined, have been improving gradually around the world for many decades, indeed for centuries. Treaties negotiated and ratified over a forty-year period beginning in the 1960s came very late in this rights improvement process, and so a question arises as to whether continued improvement in human rights outcomes *after* those treaties were ratified was caused by the treaty ratifications or by the same factors that caused human rights to improve *before* the treaties were ratified.

In asking this question, we take inspiration from a 2008 study by Daron Acemoglu, Simon Johnson, James A. Robinson, & Pierre Yared [AJRY], who asked a similar question about empirical research on the relationship between

Copyright © 2018 by Adam S. Chilton & Eric A. Posner.

This article is also available online at <http://lcp.law.duke.edu/>.

* Adam Chilton, Assistant Professor of Law and Walter Mander Research Scholar, University of Chicago Law School (adamchilton@uchicago.edu). Eric Posner, Kirkland & Ellis Distinguished Service Professor of Law and Arthur and Esther Kane Research Chair, University of Chicago Law School (eric_posner@law.uchicago.edu). We thank Matt Blackwell, Chris Fariss, Kyle Rozema, and Bartek Woda for sharing data, code, and advice. An early version of this paper circulated under the title: “Respect for Human Rights: Law and History.”

1. See, e.g., BETH A. SIMMONS, MOBILIZING FOR HUMAN RIGHTS: INTERNATIONAL LAW IN DOMESTIC POLITICS (2009); Courtenay R. Conrad & Emily Hencken Ritter, *Treaties, Tenure and Torture: The Conflicting Domestic Effects of International Law*, 75 J. POL. 397 (2013); Daniel W. Hill, *Estimating the Effects of Human Rights Treaties on State Behavior*, 72 J. POL. 1161 (2010); Yonatan Lupu, *Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements*, 67 INT’L ORG. 469 (2013a); Yonatan Lupu, *Legislative Veto Players and the Effects of International Human Rights Agreements*, 59 AM. J. POL. SCI. 578 (2015).

2. See, e.g., Oona Hathaway, *Do Human Rights Treaties Make a Difference?*, 111 YALE L. J. 1935 (2002); Linda Camp Keith, *The United Nations International Covenant on Civil and Political Rights: Does it Make a Difference in Human Rights Behavior?*, 36 J. PEACE RES. 95 (1999).

income and democracy.³ The earlier empirical research had discovered a positive correlation between national income and the existence of democracy in a country, and argued that higher income levels cause countries to become more democratic.⁴ However, AJRY found that the correlation between income and democracy is not robust to the simple inclusion of country-fixed effects or other panel regression methods that account for trends in both income and democracy. That is, after including a variable to account for the unique features of each country, the results were no longer statistically significant. Moreover, their results were precisely estimated and had confidence intervals that ruled out the possibility of all but small effects of income on democracy. Their findings thus called into question the widely accepted causal relationship between income and democracy, and suggested that the positive correlation is attributable to long-run determinants of both factors. Historical factors—which may not be known, and hence are omitted from regressions—seem to account both for the income level of a country and the country’s level of democracy.

For the question of the relationship between human rights treaties and human rights outcomes, a similar problem must be faced. The existing literature has not fully appreciated the problems that trends in treaty ratification and human rights outcomes present for causal identification. The challenge arises from two features of the data.

First, the key independent variable of interest—treaty ratification—does not show much variation over time. Countries typically ratify a treaty in a given year and then the treaty remains law thereafter. Countries almost never withdraw from a human rights treaty. Thus, the key independent variable in these studies changes at most only once per country. Moreover, countries necessarily are more likely to have ratified treaties in later years than in earlier years.

Second, countries’ human rights records also exhibit clear serial correlation over time. Countries’ human rights outcomes change slowly over a very long period of time with the treaty ratifications coming, as we noted above, late in this process.⁵

These two features of the data create problems for empirical research. The fact that both ratification and human rights outcomes are likely to have higher values in later years than earlier years creates a spurious regression problem.⁶

3. Daron Acemoglu, Simon Johnson, James A. Robinson, & Pierre Yared, *Income and Democracy*, 93 AM. ECON. REV. 808 (2008).

4. See, e.g., ROBERT A. DAHL, POLYARCHY: PARTICIPATION AND OPPOSITION (1971); SAMUEL P. HUNTINGTON, THE THIRD WAVE: DEMOCRATIZATION IN THE LATE TWENTIETH CENTURY (1991); DIETRICH RUESCHEMEYER ET AL., CAPITALIST DEVELOPMENT AND DEMOCRACY (1992); Robert J. Barro, *Determinants of Democracy*, 107 J. POL. ECON. 158 (1999).

5. See Adam Chilton & Eric A. Posner, *The Influence of History on States’ Compliance with Human Rights Obligations*, 56 VA J. INT’L L. 211 (2016); Christopher J. Fariss, *Respect for Human Rights Has Improved Over Time: Modeling the Changing Standard of Accountability*, 108 AM. POL. SCI. REV. 297 (2014); Yonatan Lupu, *Explaining Human Rights Abuses: Comparing Contemporary Factors and Historical Factors*, 56 VA J. INT’L L. 481 (2016).

6. A spurious regression problem occurs when two variables without a causal relationship are found to be related to each other either due to coincidence or an unrelated factor influencing both

Additionally, the fact that both the independent and dependent variables exhibit serial correlation can result in false rejection of the null hypothesis that laws have no effect on a given outcome.⁷ Taken together, failure to account for these features of the data may lead researchers to mistakenly conclude that treaties have a positive effect on human rights.

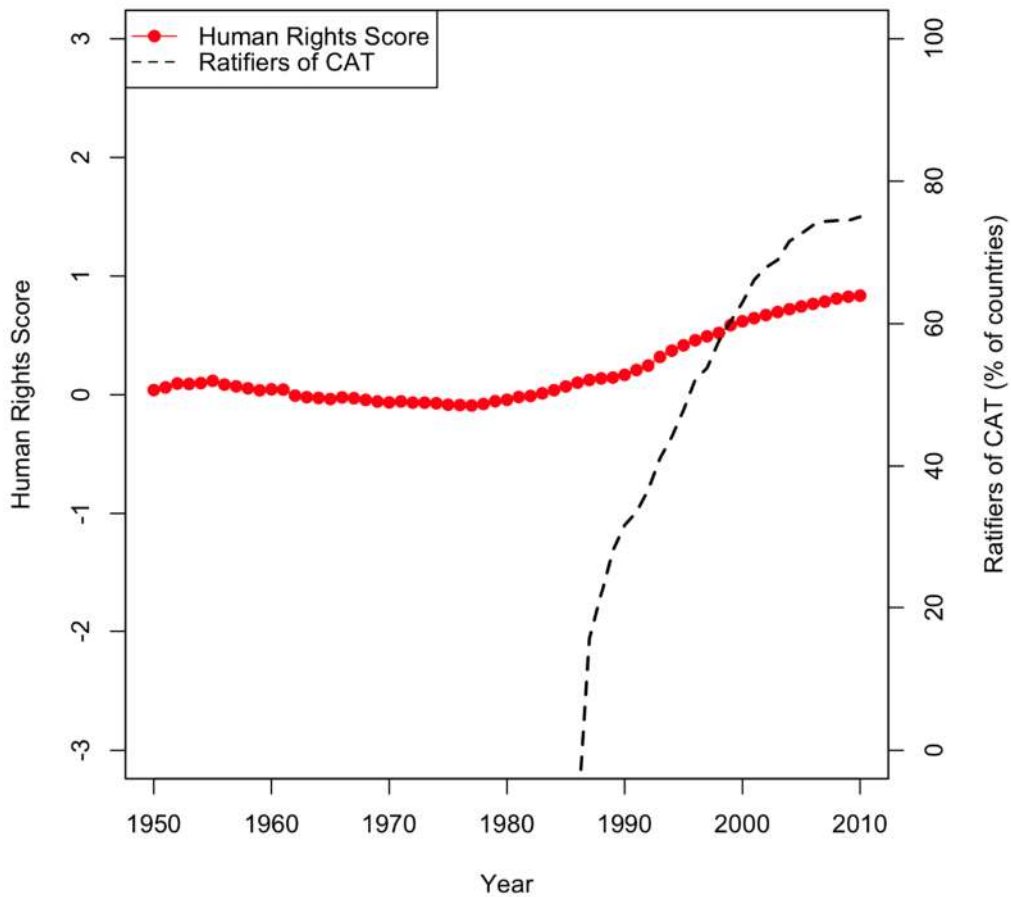


Figure 1: Human Rights Scores and Ratification of the CAT, 1950–2010

To illustrate these points, Figure 1 shows ratifications for a widely studied treaty—the Convention against Torture (CAT)—and a measure of state repression globally from 1950 to 2010. The data on state repression is from an index created by Fariss, known as the Human Rights Scores, that reports how many standard deviations above or below the global average level of repression

variables.

7. See Marianne Bertrand, Esther Dufo & Sendhil Mullainathan, *How Much Should We Trust Differences-in-Differences Estimates?*, 119 Q. J. ECON. 249 (2004).

a given country is in a given year.⁸ The index ranges from roughly -3 (worse because of more repression) to roughly 3 (better because of less repression). Figure 1 suggests that the improvement in repression, which seems to have started in the 1970s, predated ratification of the CAT in the 1980s, 1990s, and early 2000s. Although Figure 1 is merely exploratory, it puts in stark relief the question of whether treaty ratification causes improvements in outcomes; or whether, like the relationship between income and democracy studied by AJRY,⁹ the correlation between treaty ratification and improved human rights outcomes is due to historical factors omitted from the standard regressions.

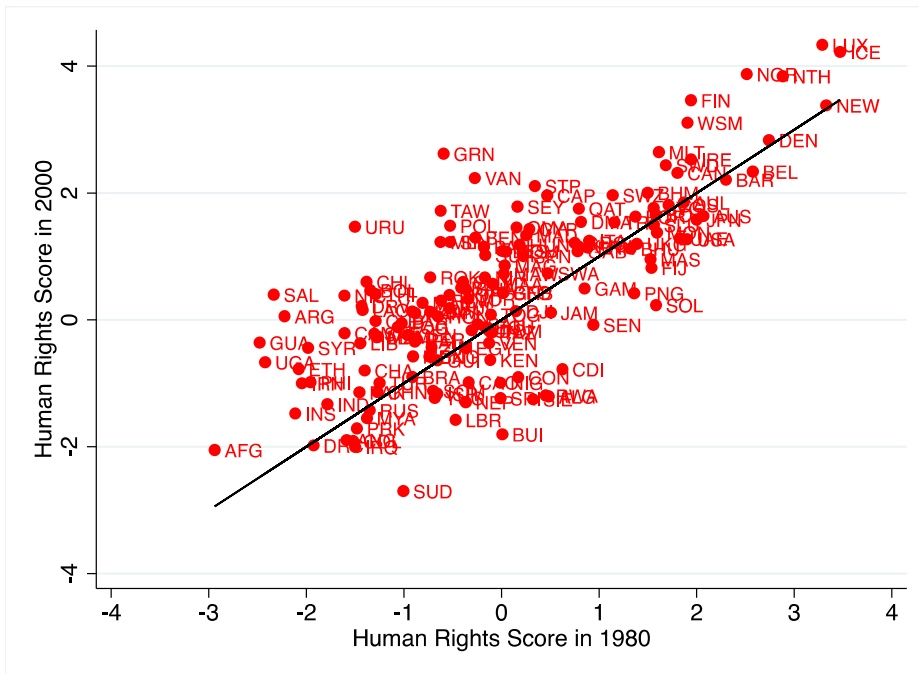


Figure 2: Human Rights Scores in 1980 vs. Human Rights Scores in 2000

Figure 2 provides another illustration. It shows the relationship between the Human Rights Score that countries received in 1980—before the CAT went into effect—and the score that countries received for the same measures in 2000. The solid line is a 45-degree line. Countries below the line had worse human rights outcomes in 2000 than 1980 for the given measure, and countries above the line had better human rights for the measure in 2000 than in 1980. As Figure 2 shows, there is a strong, positive relationship between human rights outcomes in 1980

8. See Fariss, *supra* note 5.

9. See Acemoglu et al., *supra* note 3.

and 2000. The relationship between the Human Rights Scores in 1980 and 2000 is also highly statistically significant ($p < 0.001$) and the r-squared is 0.56. In other words, countries' human rights improved over this period, but the countries' respect for human rights before the treaties went into effect is extremely predictive of the levels of respect after.

Figures 1 and 2 show the importance of patterns in ratifications and trends in human rights outcomes. Research designs that empirically compare country-year observations with treaty ratifications to country-year observations without treaty ratifications are implicitly comparing earlier country-years to later country-years. Given the gradual changes in human rights over time, this means that many research designs are prone to falsely finding positive correlations. When this possibility is combined with the large number of degrees of freedom scholars have when designing empirical tests—for example, by examining different treaties, dependent variables, and sets of countries—it is unsurprising that many papers find positive correlations between treaties and human rights.¹⁰ But, like with the relationship between income and democracy, these positive correlations may simply reflect trends in the data.

Although the possibility that trends in the data may lead to spurious correlations is a common concern for many research designs studying the effects of laws,¹¹ scholars studying international law should be particularly wary of the threat of false positives. Even when regressions present evidence of positive correlations between a variable of interest and a relevant explanatory variable, it is still possible that relationships are an artifact of “selection on the unobservable.” That is, the correlations may be due to an unobserved factor that is causing changes in both the independent and dependent variables. As Chaudoin, Hays, & Hicks showed, this problem is acute in the case of international relations and international law because the same factors that drive compliance also may cause countries to form and join international treaties or institutions.¹² Moreover, “[t]his problem . . . most likely biases empirical findings regarding the effects of institutions in a positive direction, because countries that are most likely to comply *ex ante* are also the most likely to ratify.”¹³ To illustrate this point, Chaudoin, Hays, & Hicks replicate many existing studies to demonstrate that those studies' research designs can be used to show that participation in institution can cause improvements in outcome variables that are theoretically unrelated to those institutions. Based on these results, Chaudoin,

10. See Andrew Gelman & Eric Loken, *The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There is No 'Fishing Expedition' or 'P-hacking' and the Research Hypothesis Was Posited Ahead of Time* (Nov. 14, 2013) (unpublished manuscript) (http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf) [<https://perma.cc/6STH-NHL9>].

11. See, e.g., Eric Helland & Alexander Tabarrok, *Using Placebo Laws to Test 'More Guns, Less Crime'*, 4 *ADVANCES ECON. ANALYSIS & POL.* 1 (2004); Bertrand, *supra* note 7.

12. Stephen Chaudoin, Jude Hays, & Raymond Hicks, *Do We Really Know the WTO Cures Cancer?*, 48 *BRIT. J. POL. SCI.* 903 (2018).

13. *Id.* at 903.

Hays, & Hicks argue that scholars of international institutions and law should be vigilant against reporting false positives.¹⁴

To illustrate how positive correlations between treaties and human rights may be driven by trends in ratification practices and the human rights data, we present the results of three simple empirical exercises using methods that are already common in economics and political science.

First, we use data visualization to show trends in human rights outcomes before and after ratification. Graphs similar to those widely used in event studies can help illustrate whether the trends actually emerge before a treatment begins.¹⁵ Here, these graphs raise questions about the effect of human rights treaties, as they suggest that treaty ratification does not accelerate or otherwise affect a preexisting trend in human rights outcomes.

Second, we estimate a series of regression models using panel data—in our case, data where the unit of observation is the country-year—to test the robustness of the relationship between ratification and human rights outcomes. There is a range of regression techniques that exploit the time dimension of panel data to account for unobserved-but-fixed omitted variables. Many papers in the human rights literature use one of these methods, such as the fixed-effect model or lagged dependent variable models. But it is difficult to know *ex-ante* which method is the most appropriate. As a result, the recommended best practice in economics is to estimate multiple models to test the robustness of results and to put a bound on the possible size of any effect.¹⁶ We estimate regressions using five different basic models for accounting for trends in the data for three human rights treaties. When doing so, we do not find evidence of a robust relationship between treaty ratification and various measures of human rights outcomes.

Third, we conduct placebo tests to validate the reliability of the results we generated using panel regression methods. Placebo testing can be used to determine whether regression specifications and research designs are prone to finding spurious correlations or incorrectly estimating standard errors.¹⁷ Although there are many different forms of placebo testing, the approach we use is to randomly assign countries to have “ratified” treaties and then conduct Monte Carlo simulations. This involves regressing human rights outcomes on the random treaty ratifications to see if prior estimates of the effect of human rights treaties differ from the estimates that are produced by random ratification.¹⁸ The Monte Carlo simulation results suggest that the positive and statistically

14. *Id.*

15. See, e.g., Raj Chetty & Adam Szeidl, *Consumption Commitments and Risk Preferences*, 122 Q. J. ECON. 831 (2007); Anup Malani & Julian Reif, *Interpreting Pre-Trends as Anticipation: Impact on Estimated Treatment Effects from Tort Reform*, 124 J. PUB. ECON. 1 (2015).

16. See JOSHUA D. ANGRIST & JORN-STEFFEN PISCHKE, *MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST’S COMPANION* 245–46 (2009).

17. See Robert S. Erikson, Pablo M. Pinto & Kelly T. Rader, *Randomization Tests and Multi-Level Data in U.S. State Politics*, 10 ST. POL. & POL’Y Q. 180 (2010); Bertrand et al., *supra* note 7; Chaudoin, Hays, & Hicks, *supra* note 12; Helland & Tabarrok, *supra* note 11.

18. Cf. Bertrand et al., *supra* note 7.

significant results uncovered during the initial panel regressions are within the bounds of what would be produced by random ratification. Therefore, the results of the initial panel regressions may be a falsely positive correlation

We argue that human rights researchers should use these three techniques—data visualization, multiple panel regression models, and placebo tests—in future research in order to address skepticism about the robustness of their findings.

We should mention at the outset that we do not try to replicate and debunk particular studies. Our goal is merely to illustrate problems and suggest ways to address them, and thus we use as a baseline a somewhat stylized version of existing research. The regressions presented in this paper use basic control variables and only six measures of human rights outcomes, and they do not test for conditional effects.¹⁹ More complex models may produce different results, and different research questions will call for different methods.

This paper proceeds as follows. In Part II, we briefly review the empirical literature on treaties and human rights and explain how patterns in ratification and human rights data make it difficult to test the effectiveness of human rights treaties. In Part III, we use three basic methods to show that positive correlations between human rights treaties and outcomes may be attributable to patterns in ratification and trends in human rights outcomes. Finally, in Part IV we discuss the implications of our arguments for future research.

II

BACKGROUND

A. Empirical Research on the Effect of Human Rights Treaties

The empirical literature on human rights treaties can be divided into three phases.²⁰ In the first phase, papers studying the effect of human rights treaties found that they either have null effects or negative effects on human rights outcomes.²¹ For instance, Keith studied the ratification of the UN International

19. All three decisions follow Fariss (2018), who used a similarly straightforward approach to argue that there is a positive correlation between treaties and human rights. See Christopher J. Fariss, *The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance*, 48 BRIT. J. POL. SCI. 239 (2018). In fact, we use replication data from Fariss (2018) as the starting point of our analysis. We depart from Fariss, however, by also using data from the Cingranelli-Richards Human Rights Data Project (CIRI) as a dependent variable following Hill, *supra* note 1, and Yonatan Lupu, *The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects*, 57 AM. J. POL. SCI. 912 (2013).

20. For more comprehensive reviews of this literature, see Kevin L. Cope & Cosette D. Creamer, *Disaggregating the Human Rights Treaty Regime*, 56 VA. J. INT'L L. 459 (2016); Emilie M. Hafner-Burton, *A Social Science of Human Rights*, 51 J. PEACE RES. 273 (2014); Emilie M. Hafner-Burton, *International Regimes for Human Rights*, 15 ANN. REV. POL. SCI. 265 (2012); Gregory Shaffer & Tom Ginsburg, *The Empirical Turn in International Legal Scholarship*, 106 AM. J. INT'L L. 1, 19–25 (2012); Beth A. Simmons, *Treaty Compliance and Violation*, 13 ANN. REV. POL. SCI. 273 (2010).

21. See, e.g., HEATHER SMITH-CANNOY, *INSINCERE COMMITMENTS: HUMAN RIGHTS TREATIES, ABUSIVE STATES, AND CITIZENS ACTIVISM* (2012); Emilie M. Hafner-Burton & Kiyoteru Tsutsui, *Human Rights in a Globalizing World: The Paradox of Empty Promises*, 110 AM. J. SOC. 1373 (2005); Eric Neumayer, *Do International Human Rights Treaties Improve Respect for Human Rights?*, 49 J.

Covenant on Civil and Political Rights (ICCPR) by 178 countries from 1976 to 1993 and found that ratification did not result in improvements in the four tested measures of rights protections.²² Similarly, Hathaway studied the impact of the major human rights treaties mentioned above in a sample of 166 countries over roughly forty years and found that there was “not a single treaty for which ratification seems to be reliably associated with better human rights practices and several for which it appears to be associated with worse practices.”²³ Hafner-Burton & Tsutsi also found that ratification of the six major human rights treaties does not predict government respect for human rights.²⁴

In the second phase, papers studying the effect of human rights treaties found that they have positive effects on certain human rights outcomes under certain conditions.²⁵ One of those conditions is a country’s level of democratization. Simmons studied the effect of thirteen human rights treaties and found that although seven of the agreements were associated with improved rights practices, the effects were largely confined to countries that were transitioning democracies, which are defined as not being a stable autocracy or a stable democracy throughout the post-war period.²⁶ Relatedly, von Stein found that democracies that ratify the Minimum Age Convention have lower levels of prohibited child labor.²⁷ Other scholars examined conditions other than democracy. For example, Lupu found that ratification of the ICCPR is associated with improved rights practices when there are more legislative veto players in a country,²⁸ and Cole found that ratification is associated with improved rights protestations for ratifiers with high levels of bureaucratic capacity.²⁹

Finally, in the third phase, a number of papers revisited the reliability of ways

CONFLICT RESOL. 925 (2005); Hathaway, *supra* note 2; Keith, *supra* note 2.

22. The four measures of rights protections are Freedom House Political Rights, Freedom House Civil Rights, State Department Personal Integrity Rights, and Amnesty International Personal Integrity Rights. *See* Keith, *supra* note 2.

23. Hathaway, *supra* note 2, at 1940. The six major human rights treaties are the International Convention on the Elimination of All Forms of Racial Discrimination; the International Convention on Economic, Social, and Cultural Rights; the International Convention on Civil and Political Rights; the Convention on the Elimination of All Forms of Discrimination against Women; the Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment; and the Convention on the Rights of the Child.

24. *See* Hafner-Burton & Tsutsi, *supra* note 21.

25. *See* Wade M. Cole, *Government Respect for Gendered Rights: The Effect of the Convention on the Elimination of Discrimination against Women on Women’s Rights Outcomes, 1981-2004*, 57 INT’L STUD. Q. 233 (2013); Mala Htun & S. Laurel Weldon, *The Civic Origins of Progressive Policy Change: Combating Violence Against Women in Global Perspective*, 106 AM. POL. SCI. REV. 548 (2012); Emilia Justyna Powell & Jeffrey K. Staton, *Domestic Judicial Institutions and Human Rights Treaty Violation*, 53 INT’L STUD. Q. 149 (2009); Conrad & Ritter, *supra* note 1; Hill, *supra* note 1; Lupu (2013a), *supra* note 1; Lupu, *supra* note 19; Lupu (2015), *supra* note 1; SIMMONS, *supra* note 1.

26. *See* SIMMONS, *supra* note 1.

27. *See* Jana von Stein, *Making Promises, Keeping Promises: Democracy, Ratification and Compliance in International Human Rights Law*, 46 BRIT. J. POL. SCI. 655 (2016).

28. *See* Lupu (2015), *supra* note 1.

29. *See* Cole, *supra* note 25.

to measure human rights,³⁰ and while doing so, some have tested the correlation between treaty ratifications and human rights outcomes.³¹ These papers do not try to implement strategies to identify the causal effects of treaty ratification, but instead focus on how to measure human rights outcomes. But while doing so, several have found a positive correlation between ratification and human rights outcomes.³²

During all three phases, however, there have been questions about the reliability of the methods used in this literature.³³ These criticisms have largely focused on the ability of existing methods to account for selection effects in treaty ratification³⁴ and on the quality of the data used to measure human rights outcomes.³⁵ But one problem that has not received adequate attention is the risk of false positives due to trends in treaty ratification and human rights outcomes.

30. See Christopher J. Fariss, *Are Things Really Getting Better?: How to Validate Latent Variable Models of Human Rights*, 48 BRIT. J. POL. SCI. 275 (2018b); Fariss, *supra* note 5; Fariss, *supra* note 19.

31. See Fariss, *supra* note 5; Fariss, *supra* note 19; Fariss, *supra* note 30.

32. See Daniel W. Hill & Zachary M. Jones, *An Empirical Evaluation of Explanations for State Repression*, 108 AM. POL. SCI. REV. 661 (2014); Fariss, *supra* note 5; Fariss, *supra* note 19; Fariss, *supra* note 30. It is worth noting that Hill & Jones (2014) do not focus on measuring human rights outcomes. Instead, they focus on testing the correlates of repression, which included testing the correlation between the ICCPR and the CAT on repression. Like Fariss (2014, 2018a, 2018b), however, Hill & Jones (2014) do not attempt to identify the causal effects of treaty ratification.

33. See, e.g., Adam Chilton & Dustin Tingley, *Why the Study of International Law Needs Experiments*, 52 COLUM. J. TRANSNAT'L L. 173 (2013); Ryan Goodman and Derek Jinks, *Measuring the Effects of Human Rights Treaties*, 14 EURO J. INT'L L. 171 (2003); Eric Posner, *Some Skeptical Comments on Beth Simmons's Mobilizing for Human Rights*, 44 N.Y.U. J. INT'L L. & POL. 819 (2012); Jana von Stein, *Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance*, 99 AM. POL. SCI. REV. 611 (2005); Chaudoin, Hays, & Hicks, *supra* note 12; Fariss, *supra* note 5; Hill, *supra* note 1; Lupu, *supra* note 19.

34. See, e.g., Lupu, *supra* note 19.

35. See, e.g., David Cingranelli & Mikhail Filippov, *Are Human Rights Practices Improving?*, AM. POL. SCI. REV. 1 (2018a), <https://doi.org/10.1017/S0003055418000254> [<https://perma.cc/33B5-LYNH>]; Fariss, *supra* note 5.

B. Ratification Patterns and Trends in Human Rights

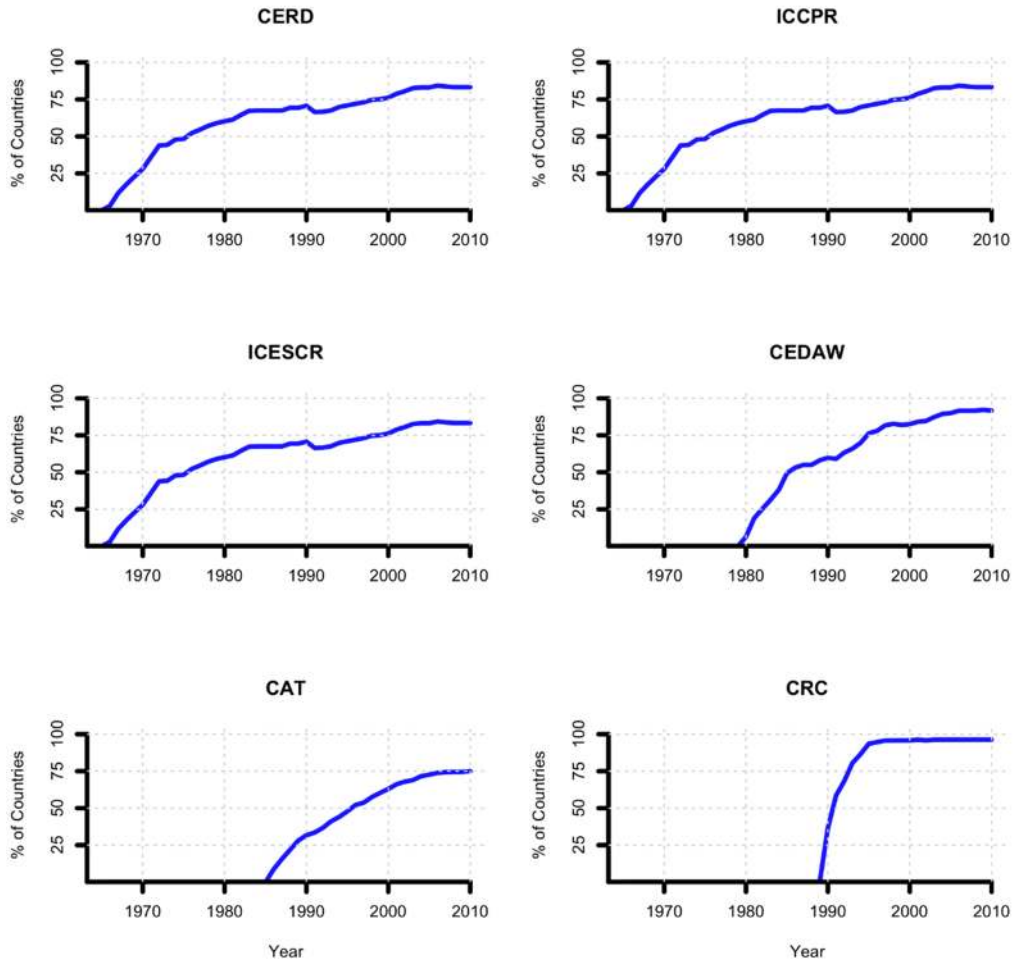


Figure 3: Ratification of Major Human Rights Treaties

An intuitive problem with testing the effects of human rights treaties is that both ratification of treaties and human rights outcomes exhibit clear trends over time. To begin, there are trends in the ratification of treaties. Figure 3 plots the ratification patterns for the six major human rights treaties. As Figure 3 shows, each of the treaties has been widely ratified in the last several decades. In the case of treaty ratifications, a country's ratification status changes at most once (from non-ratified to ratified) during the period that researchers study. Moreover, countries that ratify treaties have not unratified the treaties. Ratification is thus like a one-way switch. The result is that a given country is always a non-ratifier in earlier years and, if it decides to join the treaty, a ratifier in later years. Taken

together, this means both that the low level of changes in ratification ensures high rates of serial correlation in the key independent variable,³⁶ and that ratification is necessarily more prevalent in later years than early years.

In the case of human rights outcomes, the dependent variables used to quantify them are general measures of political and social well-being that show improvement over many decades and even centuries.³⁷ Democracy, for example, had been spreading for almost two centuries before the ICCPR, which protects political freedoms, was negotiated.³⁸ So had literacy, which is promoted by the Convention on Economic, Social, and Cultural Rights (ICESCR).³⁹ Health and political freedom outcomes for women had begun improving many decades before countries ratified the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW).⁴⁰ A further plausible cause for the improvement in many human rights outcomes is economic growth, since wealthier countries can more easily afford rights protections.⁴¹ Economic growth also goes back centuries.

Taken together, these trends in the ratification of human rights treaties and human rights improvements create two problems that would lead naïve regressions—those that simply regressed many measures of human rights (e.g., women’s economic empowerment) on treaty ratification (e.g., CEDAW)—to find an effect regardless of whether the treaty has had any effect. First, the fact that both variables exhibit trends over time creates a spurious regression problem. Second, both the independent variable (treaty ratification) and dependent variables of interest (human rights outcomes) are serially correlated, which may lead to incorrect estimation of standard errors.

First, the presence of both dependent and independent variables exhibiting trends creates a “spurious regression problem.” Although the threat posed by spurious regressions is well documented and studied in the social sciences,⁴² it has received scant attention in the human rights literature. But the problem is particularly acute in this context. The reason is that trends in human rights

36. See, e.g., David S. Abrams, *Estimating the Deterrent Effect of Incarceration Using Sentencing Enhancements*, 4 AM. ECON. J. APPLIED ECON. 32 (2012).

37. See, e.g., STEPHEN PINKER, *THE BETTER ANGELS OF OUR NATURE: WHY VIOLENCE HAS DECLINED* (2011); Max Roser, *The Short History of Global Living Conditions and Why it Matters that We Know It*, OUR WORLD IN DATA (2018), <http://ourworldindata.org/a-history-of-global-living-conditions-in-5-charts> [<https://perma.cc/95N7-B68L>].

38. See Acemoglu et al., *supra* note 3.

39. See Roser, *supra* note 37.

40. See SARAH CARMICHAEL, SELIN DILLI, & AUKE RIKPMA, *WOMEN IN GLOBAL ECONOMIC HISTORY*, in *A HISTORY OF THE GLOBAL ECONOMY FROM 1500 TO THE PRESENT* (Joerg Baten ed., 2016).

41. Of course, not all measures of human rights display positive trends. For instance, Fariss (2014) discusses how the CIRI data suggests many rights have not improved over time. See Fariss, *supra* note 5. Although there is thus some debate about whether all measures of human rights have positive trends (compare Fariss, *supra* note 19 with Cingranelli & Filippov, *supra* note 35), we are unaware of any evidence that measures of human rights are not serially correlated.

42. See, e.g., C.W.J. Granger & P. Newbold, *Spurious Regression in Econometrics*, 2 J. ECONOMETRICS. 111 (1974).

patterns are almost certainly not the same across countries. As we have argued, many long run historical factors are highly predictive of current levels of respect for rights.⁴³ Moreover, Lupu further showed that these historical factors may actually be better than contemporary factors at predicting contemporary respect for rights.⁴⁴ But the long-run historical factors that influence progress over time differ from country to country. For instance, due to events hundreds of years ago,⁴⁵ some countries started the second half of the twentieth century with good human rights records and gradually improved, while other countries started with poor records and improved more dramatically (and, of course, other countries have regressed or not improved at all during this period).

The problem is that when countries exhibit different trends, the inclusion of a simple time trend or of year-fixed effects will not solve the problem.⁴⁶ In fact, depending on the structure of the data, even country-specific time trends may not be able to fully address the spurious regression problem.⁴⁷ As a result, even research designs that took some steps to account for trends—for example, by including lagged dependent variables or linear time trends in their regressions—may still not be fully accounting for trends in human rights outcomes over time.

Second, beyond the spurious regression problem, both the independent and dependent variables exhibit serial correlations over time. The problems posed by serially correlated independent variables and serially correlated dependent variables have been well documented in economics,⁴⁸ but have received little discussion in the human rights literature. The problem is that serial correlation in both variables can lead to incorrect calculation of standard errors which leads researchers to incorrectly reject the null hypotheses that a given treatment has no effect. For example, Bertrand et al. wrote a seminal paper on how these basic problems are a threat to inference in the use of any difference-in-difference (DD) research design.⁴⁹ As Bertrand et al. explain:

While serial correlation is well understood, it has been largely ignored by researchers using DD estimation. Three factors make serial correlation an especially important issue in the DD context. First, DD estimation usually relies on fairly long-time series.

43. See Chilton & Posner, *supra* note 5 (including climate, institutional transplant, and possibly institutional lag and cultural transmissions as possible historical factors).

44. See Lupu, *supra* note 5.

45. See, e.g., Mila Versteeg, *History, Geography and Rights: A Response to Chilton and Posner*, 56 VA. J. INT'L L. 501 (2016); Acemoglu et al., *supra* note 3.

46. See Salvatore Babones, *Modeling Error in Quantitative Macro-Comparative Research*, 15 J. WORLD SYS. RES. 86 (2009).

47. As Babones explains: “Period effects, time covariates, and autoregressive error structure corrections are all inadequate to account for the kinds of time trends inherent in [Quantitative Macro-comparative research (QMCR)] data structures. The fact that QMCR variables are trended at different rates in different countries (and at different times) means that time trends must be dealt with on a country-specific basis, and sometimes in complex ways even within countries. One-size-fits-all controls of the kind typically found in QMCR completely fail to adequately adjust for the effects of trended data. As a result, the coefficients on independent variables estimated based on QMCR data are, in almost all cases, heavily biased.” *Id.* at 105.

48. See, e.g., Bertrand et al., *supra* note 7.

49. See *Id.*

Our survey of DD papers, which we discuss below, finds an average of 16.5 periods. Second, the most commonly used dependent variables in DD estimation are typically highly positively serially correlated. Third, and an intrinsic aspect of the DD model, the treatment variable [of interest] changes itself very little within a state over time. These three factors reinforce each other so that the standard error for [a variable of interest] could severely understate the standard deviation of [a variable of interest].⁵⁰

These are exactly the same issues that are present in the study of human rights treaty effectiveness: the use of panel data over long time periods, serial correlation in the dependent variables, and ratification of treaties that occurs once. Although these problems have led many economists to rigorously investigate whether the positive correlations they have found are due to serial correlation in their data,⁵¹ the same level of scrutiny of positive results is not present in the human rights literature.

This is not to say that all papers finding that human rights treaties have a positive effect on human rights outcomes were spurious results, or that all papers suffer from the problem identified in this paper. But a reading of the literature suggests that this issue has been given too little attention. Readers interested in the technical details of addressing the spurious regression problem in panel data or how to ensure their standard errors are not artificially low due to serial correlation should consult the methodological literature. We now turn to illustrating three approaches that can be easily implemented to more persuasively show that positive correlations between treaty ratifications and human rights are not simply the product of long term trends.

III

EMPIRICAL APPROACH

A. Data

For the illustrative purposes of this paper, we make a series of choices about research design that are intended to approximate the approaches most commonly used in the empirical research. First, we focus on three major human rights treaties: the ICCPR, the CEDAW, and the CAT. These three treaties have been the most widely studied by empirical researchers.⁵²

Second, we use two commonly used data sources as measures of human rights outcomes. Following Hill and Lupu, we use five measures from the Cingranelli-Richards Human Rights Data Project (CIRI) to test the ICCPR, CAT, and CEDAW.⁵³ To test the effect of the ICCPR, we use CIRI's Physical Integrity variable, which is an additive scale of four personal integrity violations: torture,

50. *See Id.*

51. *See* Helland & Tabarrok, *supra* note 11; Abrams, *supra* note 36.

52. *See* Cope & Creamer, *supra* note 20.

53. *See, e.g.,* Hill, *supra* note 1; Lupu, *supra* note 19. For more information on the CIRI data, *see* David L. Cingranelli, David L. Richards & K. Chad Clay, *Data & Documentation*, CIRI HUMAN RIGHTS DATA PROJECT (Apr. 14, 2014), <http://www.humanrightsdata.com/p/data-documentation.html> [<https://perma.cc/XQY9-POCZ>].

extrajudicial killings, political imprisonment, and disappearances. Each of these four violations is coded on an ordinal scale from 0 to 2, with 0 indicating violations were frequent and 2 indicating there were no incidents. When added together, this results in a score from 0 (most physical integrity violations) to 8 (fewest physical integrity violations). To test the effect of the CAT, we use CIRI's Torture variable, which is one of the components of the Physical Integrity variable and measured on the 0 to 2 scale described above. To test the effect of the CEDAW, we use CIRI's women's Economics Rights variable (which is coded based on factors like nondiscrimination in the workplace, and equality in hiring, promotion, and pay), CIRI's women's Social Rights variable (which is coded based on factors like equal rights to inheritance, marriage, divorce, and education), and CIRI's women's Political Rights variable (which is coded based on factors like a woman's right to vote, run for office, and petition the government). All three measures are coded on an ordinal scale from 0 to 3, with higher values associated with greater respect for rights. It is important to note that although the CIRI data has been widely used to study the effectiveness of human rights treaties,⁵⁴ it has also been criticized for failing to account for changes in reporting standards over time.⁵⁵

Additionally, following Hill & Jones,⁵⁶ we also test the effect of the ICCPR and CAT using the Human Rights Scores developed by Fariss.⁵⁷ The Human Rights Scores is a measure of repression based on 13 different variables related to human rights that includes both (relatively) objective measures—like whether a country experienced a genocide in a given year—and (relatively) subjective evaluations—like whether a country was classified as a “frequent” torturer in State Department reports.⁵⁸ The objective, events-based data provides a baseline that helps model how subjective measures have evolved over time due to changes in reporting standards. These sources are then used to create a single measure of repression for countries in a given year.⁵⁹ Although the Human Rights Scores

54. See, e.g., Hill, *supra* note 1; Lupu (2013), *supra* note 1; Lupu, *supra* note 19; Lupu (2015), *supra* note 1.

55. See Keith E. Schnakenberg & Christopher J. Fariss, *Dynamic Patterns of Human Rights Practices*, 2 POL. SCI. RES. & METHODS 1 (2014); Fariss, *supra* note 5.

56. See Hill & Jones, *supra* note 32.

57. See Fariss, *supra* note 5.

58. These 13 data variables are: (1) CIRI's extrajudicial killing measure; (2) CIRI's torture measure; (3) CIRI's political imprisonment measure; (4) CIRI's disappearance measure; (5) Hawathaway's torture measure; (6) the Ill-Treatment and Torture Data Collection Project's torture measure; (7) the Political Terror Scale (PTS) Project's state department based political terror scale; (8) PTS's Amnesty International based political terror scale; (9) Harff and Gurr's massive repressive events measure; (10) Political Instability Task Force's genocide and politicide measure; (11) Rummel's genocide and democide measure; (12) the Uppsala Conflict Data Program's government killing measure; and (13) the World Handbook of Political and Social Indicators' political execution measures. See Fariss, *supra* note 5, for more information and complete citations.

59. Because the Human Rights Scores are latent variables that are measured with uncertainty, this article follows Fariss (2014) and simulate the regressions using Human Rights Scores as a dependent variable while randomly drawing from the mean and standard deviation of each country-years Human Rights Score to produce the score for each given observation in a given year. See Fariss, *supra* note 5. We

have been used to test the effect of six major human rights treaties,⁶⁰ given the variables included in the measure, it is best understood as a measure of repression or torture.⁶¹ Also, like with the CIRI data, the Human Rights Scores have been used to study repression and human rights violations in several published papers,⁶² but there have also been criticisms about the validity of the measure based on the model specifications used to create it and concerns about extrapolating from limited data.⁶³

Third, for the regressions reported in Parts III.C and III.D, we use the same basic specifications for each of the three treaties and seven dependent variables described above. This method follows Fariss's use of control variables.⁶⁴ The reduced form regressions include only three control variables: level of democratization as measured by the Polity Score,⁶⁵ the natural log of GDP per Capita,⁶⁶ and the natural log of Population.⁶⁷ All three variables have been widely used in the human rights literature, and there is good theoretical reason to believe they are linked to both treaty ratification and human rights outcomes. We lag all right-hand side variables one year and cluster standard errors at the country level. Although the dependent variables from CIRI are ordinal, we use OLS models for interpretive ease and to allow for a consistent sample when estimating fixed effect models.⁶⁸ Finally, for any given dependent variable, we use the same set of country-year observations for all regressions. That is, we drop observations with missing values for variables used in one regression for all regressions. This ensures that any changes in results are not driven by changes in the sample. Table

thank Chris Fariss for sharing code to implement these simulations.

60. See Fariss, *supra* note 19.

61. See Adam Chilton & Mila Versteeg, *The Failure of Constitutional Torture Prohibitions*, 44 J. LEGAL STUD. 417 (2015); Fariss, *supra* note 5. Fariss (2018) uses the Human Rights Score as a dependent variable for all six major human rights treaties, and in general finds positive correlations between all the treaties and higher human rights scores. See Fariss, *supra* note 19. However, the Human Rights Scores are not good ways to measure the effectiveness of the CEDAW because they do not measure women's rights or of the ICESCR because they do not measure economic and cultural rights, for example. The fact that the Human Rights Score positively correlates with variables that theoretically should be unable to affect it (e.g. ratification of treaties unrelated to repression) is further evidence that the positive correlations are spurious. Cf. Chaudoin, Hays, & Hicks, *supra* note 12.

62. See Hill & Jones, *supra* note 32; Fariss, *supra* note 5; Fariss, *supra* note 19; Fariss, *supra* note 30; Chilton & Versteeg, *supra* note 61; Lupu, *supra* note 5.

63. See David Cingranelli & Mikhail Filippov, *Problems of Model Specification and Improper Data Extrapolation*, 48 BRIT. J. POL. SCI. 273 (2018b); Cingranelli & Filippov (2018a), *supra* note 35.

64. See Fariss, *supra* note 19.

65. See Monty Marshall, Keith Jagers & Ted R. Gurr, *Polity IV Project: Political Regime Characteristics and Transitions 1800-2010 Dataset Users' Manual* (2013), <http://www.systemicpeace.org/polityproject.html> [<https://perma.cc/6YWH-DGF6>].

66. See Kristian Skrede Gleditsch, *Expanded Trade and GDP Data*, 46 J. CONFLICT RESOL. 712 (2002).

67. See *id.*

68. Our results are substantively the same when using ordered-logit models for the CIRI dependent variables.

1 provides summary statistics for this data.

	Obs.	Mean	S.D.	Min	Max
Dependent Variables					
Human Rights Score	6,276	0.09	1.28	-2.94	4.02
Physical Integrity	3,937	4.74	2.28	0	8
Torture	4,051	0.70	0.72	0	2
Women's Econ. Rights	4,049	1.27	0.69	0	3
Women's Social Rights	3,130	1.17	0.84	0	3
Women's Political Rights	4,049	1.76	0.68	0	3
Ratification Variables					
ICCPR _{t-1}	6,276	0.53	0.50	0	1
CAT _{t-1}	6,276	0.32	0.47	0	1
CEDAW _{t-1}	6,276	0.48	0.50	0	1
Control Variables					
Polity Score _{t-1}	6,276	0.45	7.52	-10	10
Log GDP Per Capita _{t-1}	6,276	8.29	1.22	4.89	13.36
Log Population _{t-1}	6,276	9.05	1.51	5.09	14.09

Table 1: Summary Statistics

B. Data Visualization

Across the social sciences, the identification revolution has led to an increased emphasis on using research designs that can produce credible causal, rather than correlational, interpretations of data. Although this has led many researchers to use increasingly sophisticated empirical methods, it has also led to a renewed emphasis on simple data visualization.⁶⁹ One of the reasons for this trend is that non-parametric data visualizations reveal patterns in the data that are obscured by regressions. For instance, Raj Chetty, a highly influential economist, has repeatedly made event study graphs and binned scatterplots a centerpiece of his papers.⁷⁰

In our own prior joint and separate work, we have emphasized reporting graphs visualizing basic data to support our arguments about human rights.⁷¹

69. See Jonathan A. Schwabish, *An Economist's Guide to Visualizing Data*, 28 J. ECON. PERSP. 209 (2014).

70. See, e.g., Raj Chetty, John N. Friedman & Jonah E. Rockoff, *Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood*, 104 AM. ECON. REV. 2633 (2014); Chetty & Szeidl, *supra* note 15.

71. See ERIC POSNER, *THE TWILIGHT OF INTERNATIONAL HUMAN RIGHTS LAW* (2014); Adam Chilton & Mila Versteeg, *Rights Without Resources: The Impact of Constitutional Social Rights on Social Spending*, 60 J. L. & ECON. 713 (2017); Chilton & Posner, *supra* note 5; Chilton & Versteeg, *supra* note 61.

More specifically, we have reported graphs similar to those used in event study research designs showing trends in human rights outcomes before and after the ratification of treaties (or, in related work, the adoption of constitutional rights).⁷² When looking at a variety of treaty and outcome measures, these basic graphs have suggested that trends in human rights outcomes either predate the ratification of treaties or are flat after adoption.⁷³

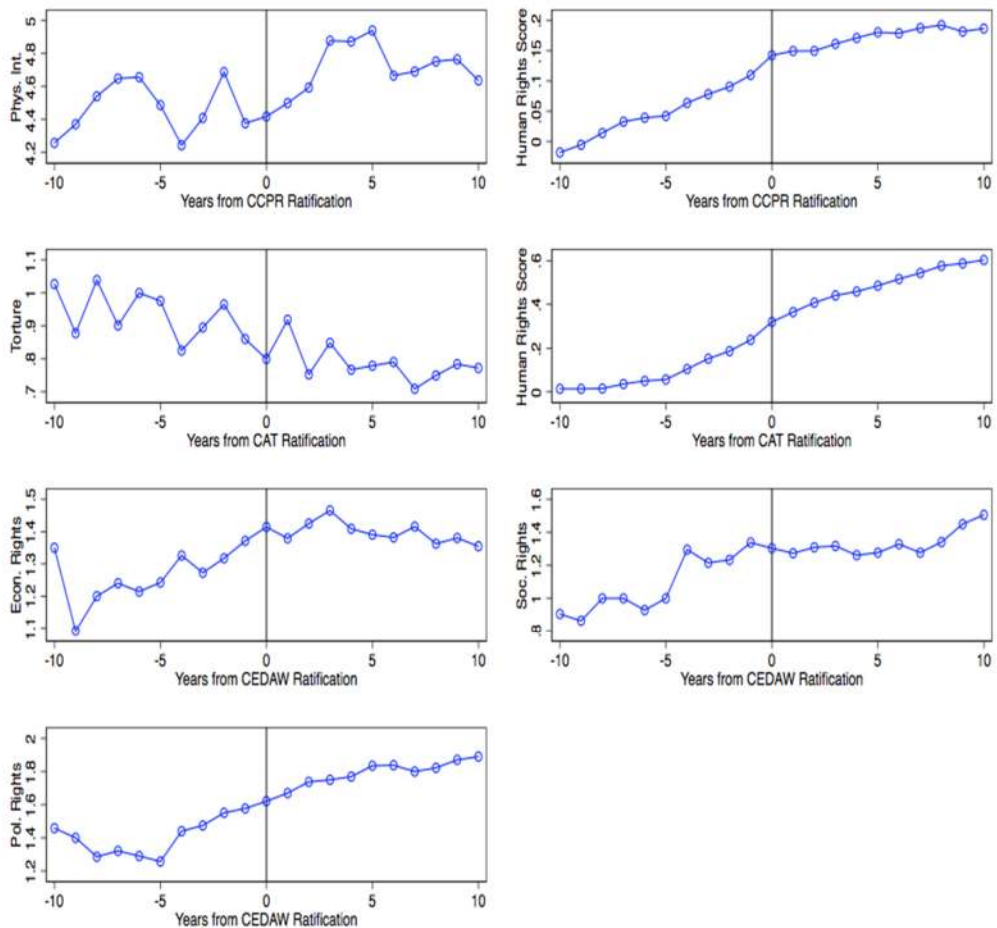


Figure 4: Event Study of Human Rights Before & After Treaty Ratification

72. See e.g., Chilton & Versteeg, *supra* note 61; Chilton & Versteeg, *supra* note 71. Acemoglu et al. (2017) use similar graphs to show changes in institutions that occur in the years before and after democratization. Daron Acemoglu et al., *Democracy Does Cause Growth*, J. POL. ECON. (forthcoming 2018), <https://doi.org/10.1086/700936> [<https://perma.cc/LYL4-JQZC>].

73. See POSNER, *supra* note 71; Chilton & Posner, *supra* note 5.

To illustrate, Figure 4 displays seven graphs that show changes in the six dependent variables listed in Part III.C for countries that ratified the ICCPR, CAT, and CEDAW in the ten years before and after ratification. The data shown in Figure 4 suggests trends towards higher human rights outcomes existed before the ratification of the relevant treaties. In fact, one of the dependent variables—the Human Rights Scores—used most recently in research suggesting a positive correlation between human rights treaties and outcomes, shows clear trends in improvement that predate the ratification of the ICCPR and the CAT.⁷⁴ This is significant, because when presented with graphs that show that trends in the dependent variable predate the treatment, “[t]he conventional diagnosis that researchers make upon observing such a pattern in the data is that the treatment was endogenous.”⁷⁵ Visual inspection of Figure 4 further suggests that the treaties likely did not accelerate the improvement in human right outcomes. Given these trends predating ratification, any analysis or regression that simply compares human rights scores before and after ratification of these treaties without accounting for the preexisting trends would find a correlation between treaty ratification and human rights outcomes—even though that correlation may simply be due to the preexisting trend.

Of course, there are limits to the evidence presented in Figure 4. To start, Figure 4 does not control for any other factors that may be associated with changes in human rights practices. Moreover, Figure 4 uses six outcome measures, but it is possible that these trends would not be present in other data.⁷⁶ Finally, Figure 4 shows a twenty-year window around ratification, but it is possible that the effects of treaties are felt later. That said, although it is possible that treaties have long-term effects, the further improvements are from ratification, the more difficult they are to credibly identify.

Given this straightforward evidence that improvements in human rights may predate treaty ratification, scholars should be cautious to ascribe causal interpretations to complex models that find positive correlations when basic non-parametric graphs of the data do not reveal changes in human rights trends after ratification. To address the concern that preexisting trends are driving their results, scholars could improve the credibility of their results by presenting data visualizations that show that positive changes in human rights outcomes occur after ratification.

C. Panel Regressions

Panel regressions have long been a staple of research on the effectiveness of human rights treaties.⁷⁷ The advantage of panel regressions is that they make it possible to account for many observable factors that influence treaty ratification

74. See Fariss, *supra* note 5; Fariss, *supra* note 19; Fariss, *supra* note 30.

75. See Malani & Reif, *supra* note 15, at 1.

76. The data we do use, however, has been used to claim a positive correlation between treaty ratification and human rights outcomes. See Fariss, *supra* note 5; Fariss, *supra* note 19.

77. See Hathaway, *supra* note 2; SIMMONS, *supra* note 1; Fariss, *supra* note 5.

and human rights outcomes, as well as unobserved factors that are fixed or relatively stable over time. These approaches thus make it possible to more adequately model the trends in human rights data that predate the emergence of treaties. But there are many different models that can be used with panel data, and it is not immediately clear which one is most appropriate for a given application. Using multiple approaches, however, reduces the possibility that any one result is model-dependent and can thus improve the panel data's reliability. Accordingly, we agree with Angrist & Pischke that the best practice should be using multiple approaches that test the robustness of results and place bounds on the size of possible effects.⁷⁸

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
ICCPR _{t-1}	0.105 (0.284)	0.376*** (0.034)					
CAT _{t-1}			-0.122* (0.062)	0.547*** (0.036)			
CEDAW _{t-1}					0.194** (0.076)	0.334*** (0.092)	0.525*** (0.086)
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Treaty Ratification on Human Rights Scores – Baseline

To illustrate how this can be done, we first present two sets of baseline results, and then use five panel regression models to test the effect of ratifying human rights treaties on human rights outcomes.⁷⁹ To begin, Table 2 presents the results of regressions on the measures of human rights outcomes discussed in Part III.A on a variable coded as 1 if a country–year observation has ratified each of three major human rights treaties (the ICCPR, CAT, and CEDAW) and 0 otherwise. These regressions do not include any control variables or fixed effects. The results for five of the seven regressions are positive and statistically significant, meaning that the regressions have a p-value of 0.1 or lower. The relationship between ratification of the CAT and CIRI's measure of torture, however, is

78. See ANGRIST & PISCHKE, *supra* note 16.

79. To be clear, we do not claim that these five models are the only ways to model the data, or even the most appropriate. There is an expansive literature on how best to model panel data, and our goal is not to rehash that literature. See, e.g., Nathaniel Beck & Jonathan N. Katz, *Modeling Dynamics in Time-Series–Cross-Section Political Economy Data*, 14 ANN. REV. POL. SCI. 331 (2011); Luke Keele & Nathan J. Kelly, *Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables*, 14 POL. ANALYSIS 186 (2006); Greg Wawro, *Estimating Dynamic Panel Data Models in Political Science*, 10 POL. ANALYSIS 25 (2002). Our goal is to simply present five different possible approaches to illustrate how researchers can use multiple models to evaluate the robustness of their results.

negative and statistically significant. These basic regressions are largely consistent with the evidence discussed in Part II.B that human rights have improved over time, and thus country–years after ratification are likely to have better human rights scores than country–years before ratification.

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
ICCPR $t-1$	-0.537*** (0.200)	0.007 (0.028)					
CAT $t-1$			-0.301*** (0.044)	0.122*** (0.030)			
CEDAW $t-1$					0.101* (0.059)	0.187** (0.073)	0.400*** (0.070)
Polity Score $t-1$	0.101*** (0.013)	0.063*** (0.002)	0.022*** (0.004)	0.060*** (0.002)	0.019*** (0.005)	0.031*** (0.007)	0.028*** (0.006)
GDP Per Capita $t-1$	0.586*** (0.078)	0.405*** (0.011)	0.191*** (0.025)	0.402*** (0.011)	0.220*** (0.031)	0.221*** (0.044)	0.034 (0.038)
Population $t-1$	-0.634*** (0.063)	-0.309*** (0.008)	-0.137*** (0.017)	-0.313*** (0.008)	-0.037** (0.019)	-0.021 (0.026)	0.024 (0.021)
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3: Treaty Ratification on Human Rights Scores – Basic Controls

Table 3 builds on these results by introducing variables to control for three factors that have been associated with human rights performance: the natural log GDP, the natural log of population, and polity scores. As previously noted, these are the same basic control variables used by Fariss when finding a positive correlation between human rights treaties and outcomes.⁸⁰ With this specification, four of the seven regressions still produce positive and statistically significant coefficients for treaty ratification. On the other hand, the ICCPR and CAT produced negative and statistically significant results that are reminiscent of the first phase of human rights research.⁸¹ These regressions, however, still do not take advantage of the panel nature of the data to account for unobserved but fixed omitted variables that may be driving trends in human rights ratification and outcomes.

80. See Fariss, *supra* note 19.

81. See, e.g., Keith, *supra* note 2; Hathaway, *supra* note 2; Hafner-Burton & Tsutsi, *supra* note 21.

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
ICCPR _{t-1}	-0.166*** (0.054)	0.003 (0.013)					
CAT _{t-1}			-0.121*** (0.021)	0.025* (0.014)			
CEDAW _{t-1}					0.023 (0.022)	0.041* (0.021)	0.077*** (0.016)
Polity Score _{t-1}	0.028*** (0.004)	0.009*** (0.001)	0.010*** (0.002)	0.008*** (0.001)	0.007*** (0.002)	0.008*** (0.002)	0.005*** (0.001)
GDP Per Capita _{t-1}	0.150*** (0.024)	0.054*** (0.006)	0.082*** (0.011)	0.054*** (0.006)	0.083*** (0.014)	0.056*** (0.013)	0.006 (0.008)
Population _{t-1}	-0.169*** (0.020)	-0.041*** (0.005)	-0.059*** (0.007)	-0.042*** (0.005)	-0.013** (0.007)	-0.006 (0.007)	0.004 (0.004)
Phys. Int. _{t-1}	0.739*** (0.018)						
HRS _{t-1}		0.873*** (0.007)		0.872*** (0.007)			
Torture _{t-1}			0.574*** (0.020)				
Econ. Rights _{t-1}					0.649*** (0.024)		
Social Rights _{t-1}						0.732*** (0.024)	
Political Rights _{t-1}							0.809*** (0.019)
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Treaty Ratification on Human Rights Scores — Lagged DV

With these baseline results established, we now turn to presenting the results of five different strategies for leveraging the panel data to account for serial correlation in the independent and depend variables over time. First, Table 4 reports the results of regressions that use the same set of control variables as Table 3, but introduce a lagged version of the dependent variable (LDV) on the right-hand side of the equation (that is, the prior year's value of the variable). The advantage of doing so is that it accounts for serial correlation in the data and captures dynamic processes in the evolution in the dependent variable.⁸² In other words, LDV models account for the possibility that current values of the dependent variable are a product of values of the dependent variable in past

82. See ANGRIST & PISCHKE, *supra* note 16.

periods and changes in other factors that influence them.⁸³ Or, in the case of human rights, the assumption is that respect for rights in the current year is a product of respect for rights in past years and other relevant changes like increases in a country's wealth. When using this model, the results in Table 4 have three regressions—the CAT's effect on Human Rights Scores, the CEDAW's effect on women's Social Rights, and the CEDAW's effect on women's Political Rights—that find a positive and statistically significant effect of treaty ratification on the relevant outcomes.

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
ICCPR _{t-1}	-0.218 (0.225)	-0.174*** (0.033)					
CAT _{t-1}			0.011 (0.052)	-0.034 (0.034)			
CEDAW _{t-1}					-0.046 (0.054)	-0.080 (0.058)	0.120* (0.066)
Polity Score _{t-1}	0.086*** (0.014)	0.034*** (0.002)	0.018*** (0.004)	0.033*** (0.002)	-0.013** (0.005)	-0.001 (0.007)	-0.001 (0.005)
GDP Per Capita _{t-1}	0.149 (0.218)	0.286*** (0.029)	0.053 (0.051)	0.296*** (0.029)	0.066 (0.067)	0.113 (0.081)	0.061 (0.076)
Population _{t-1}	-0.122 (0.508)	-0.406*** (0.076)	0.043 (0.155)	-0.404*** (0.075)	-0.342* (0.188)	-0.598** (0.281)	0.494*** (0.179)
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Treaty Ratification on Human Rights Scores – Country & Year Fixed Effects

Second, Table 5 reports the results of regressions that omit the LDVs but adds country and year fixed effects to the specifications reported in Table 3. The advantage of using country-fixed effects in the regressions is that they can account for unobserved but fixed factors that influence changes in the dependent variable.⁸⁴ In this case, the key identifying assumption of the country-fixed effects regressions is that the unobservable factors that influence the ratification of human rights treaties and human rights outcomes are time invariant (for instance, long-running historical or cultural facts about a country that do not change during the time period being studied). Although this assumption is unlikely to strictly hold, the advantages of fixed effects regressions have made them a standard part

83. See Keele & Kelly, *supra* note 79.

84. See Acemoglu et al., *supra* note 3.

of the human rights literature.⁸⁵ When using this specification, only one regression—CEDAW’s effect on women’s Political Rights—produces a positive and statistically significant coefficient for treaty ratification.

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
ICCPR _{t-1}	0.004 (0.248)	-0.053 (0.033)					
CAT _{t-1}			0.043 (0.058)	0.100*** (0.037)			
CEDAW _{t-1}					-0.031 (0.053)	-0.039 (0.052)	0.086 (0.063)
Polity Score _{t-1}	0.078*** (0.017)	0.033*** (0.003)	0.015*** (0.005)	0.032*** (0.003)	-0.002 (0.005)	-0.002 (0.006)	0.001 (0.006)
GDP Per Capita _{t-1}	-0.182 (0.261)	0.174*** (0.040)	-0.056 (0.070)	0.168*** (0.040)	-0.148** (0.067)	0.032 (0.090)	0.159* (0.093)
Population _{t-1}	3.982*** (1.481)	-0.235 (0.185)	0.324 (0.439)	-0.219 (0.184)	0.354 (0.431)	-0.376 (0.401)	0.405 (0.562)
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Spec. Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Treaty Ratification on Human Rights Scores – Country Specific Time Trends

Third, we build on the country and year fixed effects regressions by introducing country-specific linear trends, which captures trends that may be specific to each country. This specification directly models the possibility that there are also country-specific omitted factors that evolve stably over time. In the case of human rights, given the evidence that long run historical factors are predictive of modern human rights practices,⁸⁶ this is a plausible assumption. Table 6 shows that, when using this specification, only one regression—CAT’s effect on Human Rights Scores—produces a positive and statistically significant coefficient for treaty ratification.

85. See SIMMONS, *supra* note 1.

86. See Chilton & Posner, *supra* note 5; Lupu, *supra* note 5.

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
Δ ICCPR	0.313 (0.222)	0.012 (0.063)					
Δ CAT			-0.028 (0.062)	0.026 (0.058)			
Δ CEDAW					0.003 (0.054)	-0.077 (0.057)	0.009 (0.031)
Δ Polity Score	0.044*** (0.016)	0.011** (0.005)	0.007 (0.006)	0.011** (0.005)	-0.007 (0.006)	-0.003 (0.005)	0.006 (0.006)
Δ GDP Per Capita	0.173 (0.212)	0.079 (0.084)	0.092 (0.089)	0.077 (0.083)	0.012 (0.081)	-0.018 (0.079)	0.060 (0.071)
Δ Population	-1.440* (0.749)	-0.157 (0.302)	-0.137 (0.122)	-0.161 (0.302)	0.000 (0.107)	-1.032** (0.434)	0.355 (0.239)
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Treaty Ratification on Human Rights Scores — First Difference Model

Fourth, we use a first-difference model instead of a fixed effects model. The first-difference model estimates the effect of change in values of the independent variables on changes in values of the dependent variables. Like a fixed effect model, this approach addresses the problem of omitted variables; and, in a sample with two time periods, the two models are equivalent. However, when there are more than two time periods, the estimates produced by these approaches diverge, and one may be more appropriate than the other. For instance, if there is serial correlation in the error term that follows a random walk process, a first-difference model may be more appropriate than a fixed effects model.⁸⁷ The downside of applying the first-difference model here, however, is that countries only ratify treaties once, which means that the first-difference model estimates changes in human rights outcomes off of this single change in ratification. Given that any effects of treaty ratification may take time to develop, it is thus unsurprising that Table 7 reveals that there are not any statistically significant improvements in human rights outcomes after ratifying human rights treaties.

87. See MARNO VERBEEK, A GUIDE TO MODERN ECONOMETRICS (2008).

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
Ratification _{t-1}	0.118 (0.230)	-0.035 (0.067)	0.099 (0.062)	0.015 (0.068)	-0.027 (0.059)	-0.033 (0.057)	0.066 (0.064)
Ratification _{t-2}	-0.119 (0.217)	-0.036 (0.089)	-0.112* (0.057)	-0.006 (0.090)	-0.014 (0.049)	-0.035 (0.057)	0.020 (0.051)
Ratification _{t-3}	-0.228 (0.170)	-0.014 (0.089)	0.070 (0.047)	-0.023 (0.089)	0.037 (0.043)	-0.028 (0.048)	0.009 (0.031)
Ratification _{t-4}	0.123 (0.177)	-0.023 (0.088)	-0.058 (0.052)	-0.013 (0.089)	-0.054 (0.042)	-0.025 (0.042)	0.051 (0.033)
Ratification _{t-5}	-0.261* (0.146)	-0.153** (0.065)	-0.026 (0.040)	-0.035 (0.067)	-0.020 (0.049)	0.020 (0.049)	0.027 (0.048)
Polity Score _{t-1}	0.087*** (0.014)	0.034*** (0.002)	0.018*** (0.004)	0.033*** (0.003)	-0.013** (0.005)	-0.001 (0.007)	-0.002 (0.005)
GDP Per Capita _{t-1}	0.157 (0.218)	0.282*** (0.040)	0.057 (0.051)	0.298*** (0.029)	0.065 (0.067)	0.113 (0.081)	0.061 (0.076)
Population _{t-1}	-0.071 (0.506)	-0.438*** (0.076)	0.023 (0.154)	-0.416*** (0.076)	-0.338* (0.187)	-0.595*** (0.280)	0.486*** (0.180)
Cumulative Effect	-0.367 (0.262)	-0.261*** (0.037)	-0.026 (0.057)	-0.062 (0.040)	-0.077 (0.066)	-0.101 (0.080)	0.173 (0.081)
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,937	6,276	4,051	6,276	4,049	3,130	4,049

Robust standard errors clustered on country in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Treaty Ratification on Human Rights Scores – Distributed Lag Model

Fifth, since changes in human rights outcomes from treaty ratification may take more than one year to appear, we use a distributed lag model that allows for the possibility that the effect of treaty ratification may take time to develop.⁸⁸ We specifically include five lagged values of the treaty ratification variable. This model then estimates the cumulative effect of these five lagged variables on Human Rights Scores. When using this approach, the effects of treaty ratification again do not suggest a statistically significant, positive correlation between human rights ratification and improved human rights, as Table 8 shows.

88. See Acemoglu et al., *supra* note 72.

	ICCPR		CAT		CEDAW		
	Phys. Int.	HRS	Torture	HRS	Econ. Rights	Social Rights	Political Rights
Baseline (Table 2)	0	+++	-	+++	++	+++	+++
Basic Controls (Table 3)	---	0	---	+++	+	++	+++
Lagged DV (Table 4)	---	0	---	+	0	+	+++
Country & Year F.E. (Table 5)	0	---	0	0	0	0	+
Country Spec. Trends (Table 6)	0	0	0	+++	0	0	0
First Difference (Table 7)	0	0	0	0	0	0	0
Distributed Lag (Table 8)	0	---	0	0	0	0	0

0 = result is not statistically significant
 --- p<0.01, -- p<0.05, - p<0.1
 +++ p<0.01, ++ p<0.05, + p<0.1

Table 9: Treaty Ratification on Human Rights Score – Summary of Results

Taken together, the results in Tables 2 through 8 suggest that ratifications of the ICCPR, CAT, and CEDAW have had limited or no impact on human rights outcomes. Table 9 summarizes these results by reporting whether each model produced null results, statistically significant positive results, or statistically significant negative results. Although we used the same set of observations for every regression using a given dependent variable, not a single one of the treaties has consistently positive statistically significant results across all seven specifications. Additionally, for the regressions that account for the panel nature of the data—that is, the regressions reported in Tables 4 through 8—only five of the thirty-five specifications produced statistically significant and positive results. This therefore suggests that any evidence of a positive correlation between a given treaty and outcome of interest may not be robust to alternative ways of accounting for trends in the data over time. Moreover, as discussed in Part II.B, the statistically significant coefficients in even these regressions may be due to serial correlation in the ratifications and outcomes data.

Finally, it is important to note that there are many other methods that can be used to leverage this panel data to study the effect of human rights treaties, and we do not mean to imply that these methods should be given primacy.

D. Placebo Testing

Even when using panel data to account for unobservable factors, one can generate false positives. Chaudoin, Hays, & Hicks replicated a series of published international relations papers while including a new dummy variable for whether the country had joined the World Trade Organization (WTO), and then tested whether WTO membership had a statistically significant relationship with theoretically unrelated dependent variables from the original paper.⁸⁹ They found that these replications produced statistically significant results at a rate dramatically higher than chance. As noted in the introduction, the authors conclude that this illustrates the need for the international law literature to be particularly vigilant against the threat of false positives.

The approach that Chaudoin, Hays, & Hicks use is a form of placebo testing.⁹⁰ Placebo testing is a way of evaluating whether a research design produces results that it should not. These “false” results can either be finding that explanatory variables have effects on outcome variables that they should not be able to affect⁹¹ or that explanatory variables have incorrectly estimated standard errors.⁹² One variant of this approach is to construct a series of placebo laws that are randomly generated, and to use Monte Carlo simulations to estimate regressions using these placebo laws in place of the actual laws.⁹³ The coefficient obtained from the regression using real laws can then be compared to the distribution of coefficients obtained from the placebo laws. If the initial coefficient is not in the 10 percent most extreme coefficients, the researcher cannot “rule out the possibility that the effect . . . is just random noise.”⁹⁴

This method can be used in the case of human rights to test whether regressions are picking up positive results without adequately accounting for patterns in the data. To explore this possibility, we randomly assigned each country a ratification year that was any year after the treaty opened for signature. For example, although Albania ratified the CEDAW in 1994 and Armenia ratified the CEDAW in 1993, the first simulation may randomly assign Albania’s ratification to the year 2000 and Armenia’s ratification to the year 1987. Notably, we randomly assigned the same distribution of ratification years as the actual ratification of the treaty. In other words, if ten countries ratified the CEDAW in 1993 and nine countries ratified the CEDAW in 1994, our simulations would

89. See Chaudoin, Hays, & Hicks, *supra* note 12.

90. *Id.*

91. See, e.g., Chetty, Friedman, & Rockoff, *supra* note 70.

92. See, e.g., Bertrand et al., *supra* note 7.

93. See Bertrand et al., *supra* note 7; Helland & Tabarrok, *supra* note 11; Erikson, Pinto, & Radar, *supra* note 17.

94. See Erikson, Pinto & Radar, *supra* note 17, at 190. It is important to note that other test statistics can also be calculated using this approach. For instance, Erikson, Pinto, & Rader (2010) record the coefficients, z-values, and significance levels from each simulation. For simplicity, we focus on the distribution of coefficients of interest from the placebo laws test. Erikson, Pinto, & Rader *supra* note 17, at 188 (“The best test . . . is . . . [f]or each coefficient, we can determine the portion of the random draws within 0.05 (or less) at each tail and observe whether the observed coefficient falls within these bounds.”).

assign ten randomly selected countries to have ratified the treaty in 1993 and nine countries to have ratified the treaty in 1994. After assigning a ratification year, we generated a “Placebo Ratification” that codes the country as 0 in all years before the randomly generated year, and 1 in that year and all years after.⁹⁵ We next estimated the regressions for the five specifications from Part III.C that produced statistically significant, positive results when using panel regression techniques (these are the specifications from Tables 4 through 8 that are shown as having “+” or “+++” in Table 9). We saved the results for the Placebo Ratification variable from these regressions and then repeated this process 999 more times.

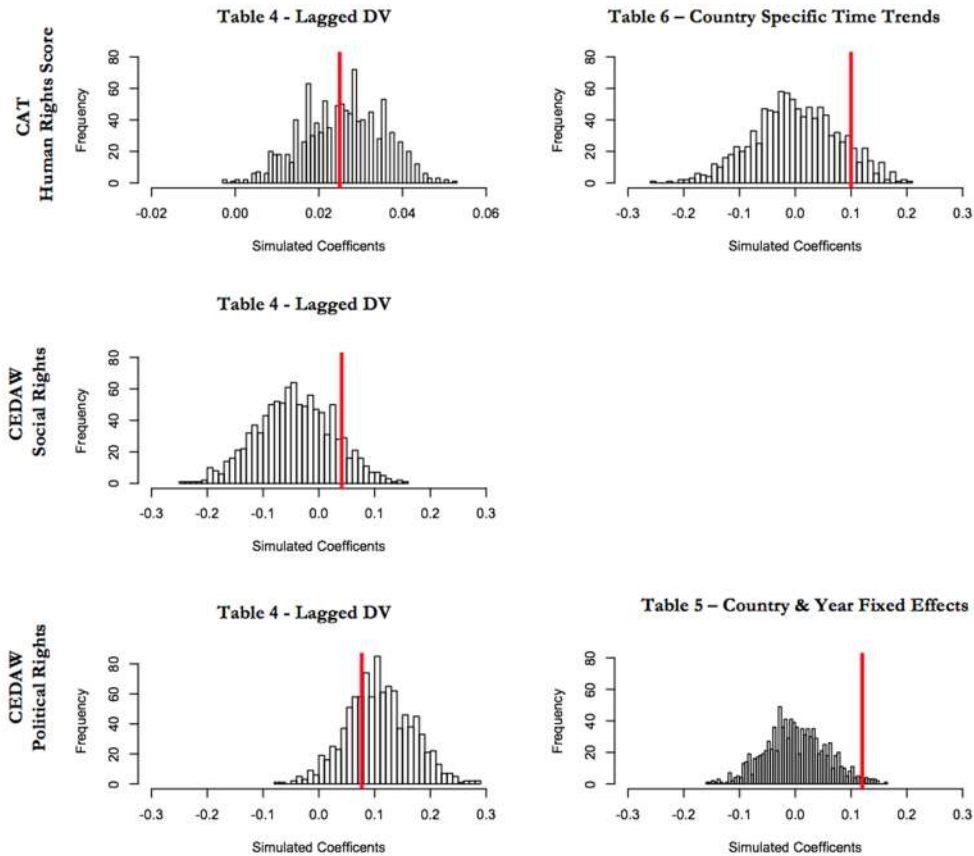


Figure 5: Distribution of P-Values from “Placebo Ratification” Simulations

95. See Helland & Tabarrok, *supra* note 11.

Figure 5 graphs the distribution of coefficients produced for each of these five specifications when using a Placebo Ratification as the key independent variable. For each distribution, the solid vertical line reports the coefficients from the regressions using the actual ratification variables that were produced in the relevant regressions in Tables 4 through 8. As Figure 5 shows, the coefficients for four of the regressions are not in the 10 percent most extreme coefficients that were generated using random ratifications, and thus the placebo test cannot rule out the possibility that the positive effects are just random noise. The one exception is the regression of women's political rights on CEDAW. The coefficient from this regression is in the top 10 percent of most extreme coefficients from the placebo ratification simulations (although it was not in the top 5 percent of most extreme coefficients). In other words, after using the placebo laws approach from Bertrand et al.,⁹⁶ one result remained positive and significant at the 0.1 level.

IV

CONCLUSION

The failure to account for trends in ratification and human rights outcomes has produced overly optimistic estimates of the effect of treaty ratification on human rights outcomes. To be clear, to say that an apparent trend explains regression results is more precisely to say that there is an omitted variable, and that while we cannot identify the omitted variable, we can show evidence that the regression results arise from an omitted variable that changes in time with the dependent variable, rather than an independent variable of interest in the regression.

There are other problems with the literature. First, the literature suffers from serious multiple-hypotheses testing problems.⁹⁷ After an initial wave of papers failed to find a positive correlation between treaties and human rights, a second wave of papers found a positive correlation under certain conditions, for specific countries, for specific treaties, using specific models. Given that researchers have many degrees of freedom when looking for patterns, scholars should be cautious about these conditional findings—however plausible the theories provided to motivate them may be.

Second, it is a matter of concern that more papers have found positive effects for the treaties where the conditions for causal inference are worse than for the treaties where the conditions for causal inference are better. The CEDAW is the treaty that the most papers have found to have a positive association with human rights practices. As previously shown, however, the best data appears to suggest that there has been slow and steady progress in women's rights over many decades, indeed centuries, and nearly all countries ratified the CEDAW in a period of a few years. As a result, research studying the effect of the CEDAW is

96. See Bertrand et al., *supra* note 7.

97. See Gelman & Loken, *supra* note 10.

essentially comparing countries' respect for women's rights before the 1980s to after the 1980s. Given the secular improvements over time, it is thus unsurprising that these papers find positive correlations. In contrast, the ICCPR was ratified slowly over a period of decades, but fewer papers have found the treaty to have a positive effect.

Third, scholars studying the effect of treaties on human rights should take note of the emerging literature on constitutional rights and human rights. A number of papers study how the inclusion of specific rights in constitutions affects human rights outcomes.⁹⁸ These papers have studied some of the same outcomes as the literature on human rights treaties, including the effects of prohibitions on torture, promises of freedom of expression, or guarantees of socioeconomic rights. However, with a few exceptions, this literature has not found that individual constitutional rights are associated with improvements in human rights outcomes. This should be a matter of concern to human rights researchers because there are strong theoretical arguments that constitutions are more likely to affect government behavior than treaties are, and that constitutional rights are methodologically better suited to testing because they are not ratified in a narrow window of time. Again, this is not dispositive. But given that the constitutional right to torture has been shown to have no correlation with rates of torture,⁹⁹ one should be skeptical of results suggesting that the CAT is having its intended effect.

Finally, even if treaties are associated with improved human rights practices, they likely explain a very small amount of the variance in human right practices.¹⁰⁰ In prior research, we argued that variables used in development economics to explain long running trends in growth are strong predictors of contemporary human rights outcomes.¹⁰¹ In a response to our project, Lupu found evidence that the historical variables we examined are better predictors of contemporary human rights outcomes than contemporary variables are.¹⁰² If these preliminary findings are correct, and more of the variance in contemporary patterns with respect for human rights can be explained by variables that are hundreds of years old, then it is hard to believe that a relatively minor shock like signing an international treaty could do much to change those trends.

98. See Adam Chilton & Milea Versteeg, *Do Constitutional Rights Make a Difference?*, 60 AM. J. POL. SCI. 575 (2016); Chilton & Versteeg, *supra* note 61; Chilton & Versteeg, *supra* note 71.

99. See Chilton & Versteeg, *supra* note 61.

100. See Hill & Jones, *supra* note 32.

101. See Chilton & Posner, *supra* note 5.

102. See Lupu, *supra* note 5.