# After Ratification: A Causal Mediation Analysis of International Human Rights Treaty


**Sinh Nguyen**
Purdue University
Department of Political Science
Email: `nguyens@purdue.edu`


## 1 Motivation

- Causal mediation of human rights treaty effect with multiple mediators.
- Roadmap: "define first, identify second, estimate last." Define in counterfactual language, identify in causal graphs, estimate with machine learning-based estimators.
- Varying causal assumptions for identification.
- Parametric regression-based estimator *vs.* machine learning-based inverse probability of treatment-weighted (IPTW) estimator.

## 2 Theory

- Treaty ratification influences human rights conditions through multiple causal pathways:
  - Directly (normative persuasion and emulation).
  - Indirectly through (1) domestic electoral accountability; (2) domestic legislative agenda-setting; (3) domestic judicial enforcement; (4) international NGOs mobilizing.
- How much does ratification of the Convention against Torture ($A$) change human rights conditions ($Y$) directly and indirectly (through $M_1$ to $M_4$) given the confounders ($W$)?

## 3 Formulation

- Structural and graphical causal models to represent the data-generating process from which $n$ observations are independently and identically sampled $O = (W, A, M_1, \ldots, M_4, Y) \sim P_O$.
- Causal quantities: $E[Y_{1,M_1}]$, $E[Y_{0,M_0}]$, and $E[Y_{1,M_0}]$.
- Causal parameters: TE $= E[Y_{1,M_1} - Y_{0,M_0}] = E[Y_{1,M_1} - Y_{1,M_0}] + E[Y_{1,M_0} - Y_{0,M_0}] =$ NIE+NDE

## 4 Identification

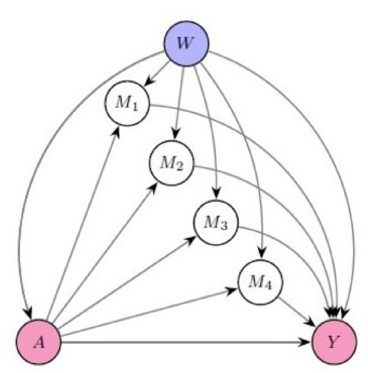


**Figure 1:** A causal DAG representing a causal story about CAT ratification $A$, mediators/mechanisms of influence $M_1$ to $M_4$, and human rights outcome $Y$. Latent factors $U$'s are assumed to be mutually independent and are not represented in the causal graph. All mediators are assumed to be conditionally causally independent.

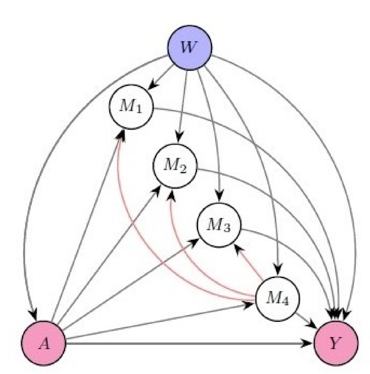


**Figure 2:** A causal DAG representing a causal story about CAT ratification, mediators, and human rights outcomes in case of non-independence among mediators. Latent factors $U$'s are assumed to be mutually independent and are not represented in the causal graph. Mediators are assumed causal dependent and considered as jointly mediating the causal effect of ratification.

- Identification conditions for TE: (1) $W_1$ leaves open causal paths from $A$ to $Y$; (2) $W_1$ blocks backdoor paths from $A$ to $Y$; (3) $W_1$ does not create spurious paths involving a collider or a descendant of a collider.
- Additional conditions for NDE and NIE: (1); $W_2$ blocks backdoor paths from $M$ to $Y$ that do not go through $A$; (2) $W_2$ blocks backdoor paths from $A$ to $M$.
- Separate sets $W_1$ and $W_2$ possible — more flexible. In practice, one sufficient set $W$.
- Causal independence among mediators: counterfactuals computable from observed conditional probability.
  - $E[Y_{1,M_1}] = E_W[Y|A=1, W=w]$
  - $E[Y_{0,M_0}] = E_W[Y|A=0, W=w]$
  - $E[Y_{1,M_0}] = E_{M,W}[Y|A=1, M=m, W=w]P(M=m|A=0, W=w)$
- Causal dependence among mediators: counterfactuals generally non-computable. TE and joint NIE still computable.

## 5 Estimation

- Observed joint distribution $P_n$ of $n = 3,992$ observations from 184 countries (1992 – 2013).

**Table 1:** Model variables

| *Sets* | *Variables and References* |
|---|---|
| $W$ | Ratification rules measured by Simmons (2009)<br>Domestic legal traditions (Mitchell, Ring and Spellman 2013) measured by La Porta, Lopez-de Silanes and Shleifer (2008)<br>Electoral rules (Cingranelli and Filippov 2010) measured by Cruz and Scartascini (2016) and Simmons (2009)<br>Treaty Commitment Propensity Lupu (2014) measured by Lupu (2014)<br>Gross domestic product (GDP) per capita (Hafner-Burton and Tsutsui 2007)<br>Participation in international trade (Hafner-Burton 2013)<br>Population size (Hafner-Burton and Tsutsui 2007)<br>Regime types (Hathaway 2007; Chapman and Chaudoin 2013; Neumayer 2007) measured by Polity IV (Marshall Monty, Keith and Robert 2016).<br>Regime durability (Goodliffe and Hawkins 2006) measured by Polity IV (Marshall Monty, Keith and Robert 2016).<br>Freedom of the press (Conrad and Moore 2010) measured by Freedom House<br>Involvement in international or domestic conflicts (Chapman and Chaudoin 2013) measured by Themnér (2014)<br>Region indicators measured by United Nations Regional Groups. |
| $A$ | Ratification of the CAT |
| $M$ | $M_1$:Electoral accountability (Dai 2005) measured by Government Vote Share (Beck et al. 2001)<br>$M_2$: Legislative agenda setting (Lupu 2015) measured by Political Constraint Index (Henisz 2002)<br>$M_3$: Judicial enforcement (Powell and Staton 2009; Conrad 2013) measured by Latent Judicial Independence (Linzer and Staton 2015) or by the Rule of Law measure (Kaufmann, Kraay and Mastruzzi 2011)<br>$M_4$: Mobilization (Murdie and Davis 2012; Simmons 2009) measured by International Non-governmental Organizations from (Lupu 2015) |
| $Y$ | Human Rights Protection Scores (Fariss 2014) |

**Figure 1:** Model variables

- Linear models of outcome and mediators (joint mediators with causal dependence) using bootstrap-based inference and linear models (individual mediators with causal independence) using simulation-based inference.
- Estimates are statistically insignificant and non-distinguishable from zero.

**Table 2:** Causal mediated effects of CAT ratification estimated using linear parametric models with bootstrap SE.

| Effects | Mean | SE | Lower | Upper |
|---|---|---|---|---|
| Natural direct effect | −0.036 | 0.026 | −0.088 | 0.015 |
| Natural indirect effect | −0.017 | 0.014 | −0.045 | 0.012 |
| Total effect | −0.053 | 0.029 | −0.110 | 0.004 |

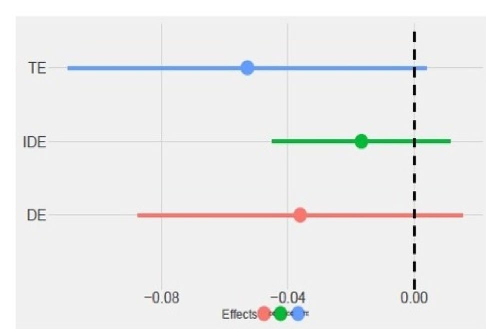


**Figure 3:** Causal mediated effects of CAT ratification on human rights outcome measured in Human Rights Scores (0 – 1 scale), 1992 – 2013. All mediators are considered jointly, that is, they simultaneously take on their natural values under either ratification or non-ratification.

**Table 3:** Natural indirect effects of CAT ratification estimated using linear parametric models with simulation-based SE. Mediators are considered individually and successively.

| Mediator | Mean | SE | Lower | Upper |
|---|---|---|---|---|
| Electoral | 0.0001 | 0.0002 | −0.0002 | 0.0004 |
| Legislative | −0.0001 | 0.0001 | −0.0003 | 0.0001 |
| Judicial | −0.0024 | 0.001 | −0.005 | 0.00004 |
| Mobilizing | −0.0001 | 0.0004 | −0.0008 | 0.0007 |

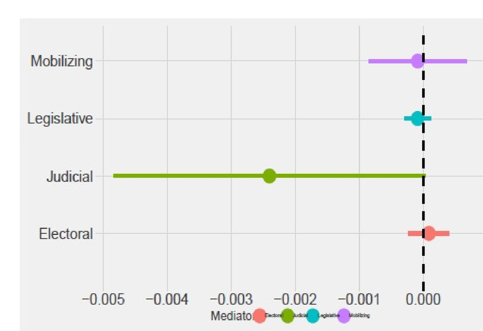


**Figure 4:** Causal mediated effects of CAT ratification on human rights outcome measured in Human Rights Scores (0 – 1 scale), 1992 – 2013. All mediators are considered individually and successively.

- Parametric models *vs.* machine learning algorithms: unknown true predictive function $Y = f(A, M, W)$; least square loss function $E[Y - Q(A, M, W)]^2$.
- Parametric models fare worse. Super Learner has the best performance, more closely approximating the true function. Flexible tools exist (e.g., `mediation` package), but still require parametric specification. Super Learner automates choices with better performance.

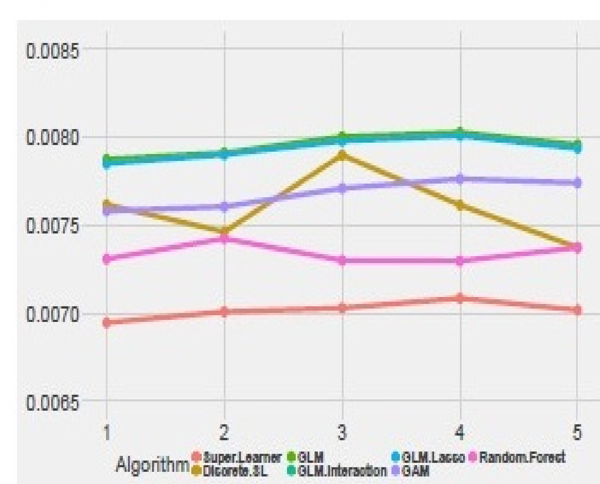


**Figure 5:** 20-fold cross-validated average risk estimates of predicting human rights outcome (measured in Human Rights Scores on 0 – 1 scale, 1992 – 2013) by seven algorithms (Ensemble Super Learner, Discrete Super Learner, Random Forest, GAM, GLM Lasso, GLM) across five imputed datasets. Cross-validated risks for GLM with two-way interaction are so high they have to be cropped out of Figure 5 for ease of presentation.

- Super Learner-based IPTW: (1) avoids modeling multiple (continuous) mediators, less computationally expensive; (2) uses Super Learner to predict weights and outcome values; (3) uses stabilized weights.
  - Compute $E[Y_{1,M_1}] = E_W[Y|A=1, W=w]$ and $E[Y_{0,M_0}] = E_W[Y|A=0, W=w]$: mean outcome values among observations with $A=1$ and $A=0$ and given SL-predicted stabilized weights $\frac{P(A=1)}{P(A=1|W)}$ and $\frac{P(A=0)}{P(A=0|W)}$, respectively.
  - Compute $E[Y_{1,M_0}] = E_{M,W}[Y|A=1, M=m, W=w]P(M=m|A=0, W=w)$: mean Super Learner-predicted outcome values among observations with $A=0$ (using their corresponding values of mediators), but fix treatment value at $A=1$ and then given Super Learner-predicted stabilized weights $\frac{P(A=1)}{P(A=1|W)}$.
  - Assumption of causal dependence among mediators only permits joint modeling of mediators. Unable to tease out portion mediated by individual mediators.
  - Natural direct effect and (joint) natural indirect effect both statistically and substantively significant.

**Table 6:** Super Learner-based estimates of natural direct and indirect effects of CAT ratification on human rights outcome (measured in Human Rights Protection Scores on a 0 – 1 scale, 1992 – 2013)

| | Mean | SE | Lower | Upper | Effects |
|---|---|---|---|---|---|
| TE | 0.158 | 0.007 | 0.145 | 0.171 |
| DE | 0.113 | 0.002 | 0.109 | 0.116 |
| $\text{IE}_{\text{joint}}$ | 0.045 | 0.008 | 0.029 | 0.061 |

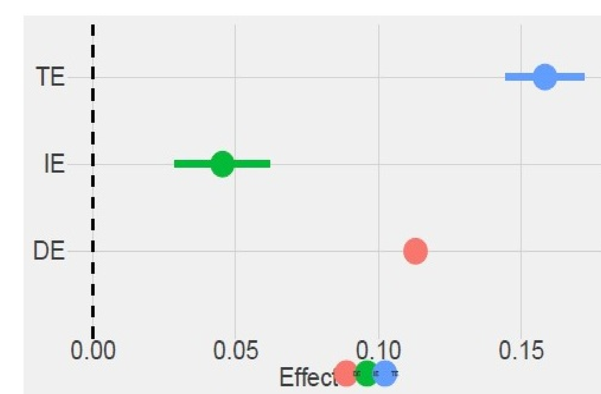


**Figure 6:** Super Learner-based estimates of natural direct and indirect effects of CAT ratification.

- Informally, $E[Y_{1,M_1}] - E[Y_{1,M_0}]$ is about how much a change in mediator due to a change in treatment will impact the outcome. $E[M|do(A=1)] - E[M|do(A=0)]$, causal effect of $A$ on each $M$, and $E[Y|do(M=1)] - E[Y|do(M=0)]$, causal effect of each $M$ on $Y$, might give some hint about the effectiveness of the legislative mechanism.

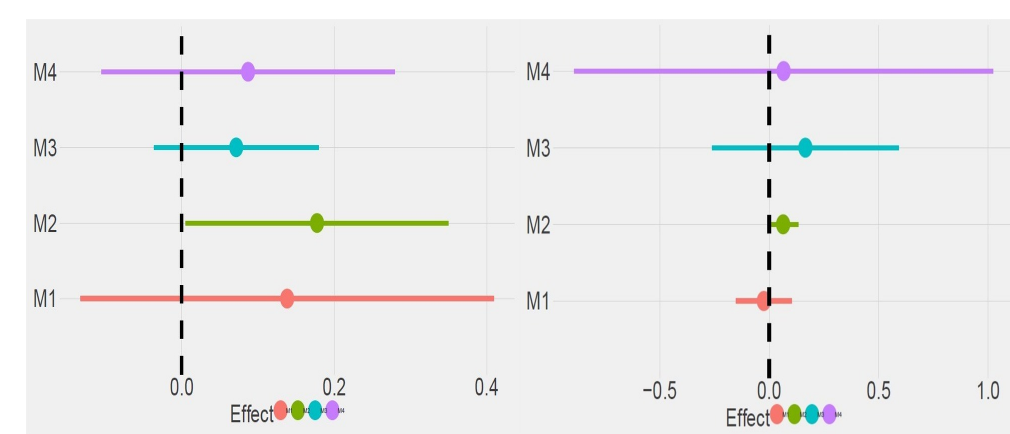


**Figure 2:** Left: causal effect of $A$ on each $M$ ($M$ on $0-1$ scale). Right: causal effect of each $M$ dichotomized at empirical mean on $Y$ ($Y$ on $0-1$ scale). M1: electoral mechanism; M2: legislative mechanism; M3: judicial mechanism; M4: international NGOs mobilizing. Identification based on causal graphs with causal dependence among mediators. Estimation uses Super Learner-based targeted maximum likelihood estimation.

## 6 Conclusion

- Further empirical analyses are needed to keep up with theoretical developments in international human rights research.
- Positive impact by treaty ratification, both directly and indirectly; particularly the direct effect of normative persuasion and possibly the indirect effect through legislative mechanism.
- Combination of recent developments in causal inference literature and machine learning research could be especially fruitful.