

Notes and Comments

Problems of Model Specification and Improper Data Extrapolation

DAVID CINGRANELLI AND MIKHAIL FILIPPOV*

Keywords: human rights scores; human rights data; democracy; model misspecification; data extrapolation; dynamic latent human rights scores

We write in response to Christopher J. Fariss, ‘The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance’. Fariss (2016) claims that the ‘standards of accountability’ of human rights violations have changed over time, and that these changes have created a systematic negative bias in standard measures of human rights. According to Fariss, this bias could be corrected by using a version of a Dynamic Item Response Model (IRM). His article purports to demonstrate the advantages of this type of data correction. He claims to show that ratifying human rights treaties really matters for human rights. However, this finding contradicts the findings of most previous research.

To demonstrate the usefulness of his measurement approach, Fariss calculates two new measures of human rights, ‘corrected’ and ‘uncorrected’ human rights scores. corrected scores assume changing standards of accountability in the records of human rights violations, and uncorrected scores are based on the conventional assumptions of IRM. The difference between the results obtained from ordinary least squares models using corrected and uncorrected scores serves as his primary evidence that the changing standards of accountability in uncorrected human rights measurement are a real problem.

However, the differences in the estimations reported in the article are very small, and are mainly due to two factors: the failure to control for the level of democracy and improper data extrapolation prior to 1981. In fact, Human Rights Treaty Ratification is not statistically related to corrected scores when controlling for the level of democracy. Taking into account changes in the level of democracy, we see no positive trend in the corrected scores, and thus no evidence of either ‘changing standards of accountability’ or an improving trend in human rights. Moreover, all differences in the results disappear when the analysis uses only data after 1980 (the period of the Cingranelli-Richards (CIRI) and Political Terror Scale (PTS) data availability). Therefore, we urge others who use the corrected scores to (1) control for the level of democracy and (2) include a separate analysis of the data after 1980 to make sure that their findings still hold.

We do not agree that the findings show that ratifying human rights treaties improves human rights performance. The findings supporting this conclusion are the artefact of the positive time trend for democracy, in parallel with the significant increase in human rights treaty ratifications (see Figure 1A of the Appendix). There are good theoretical reasons to expect that democracies would have better human rights records, and the results of *all* previous research have supported that expectation. For example, see the recent review of this literature and the conclusions of Hill and Jones.¹ Thus models explaining variation in human rights practices due to treaty ratification are misspecified unless they control for the level of democracy.

Fariss includes the results of better-specified models that incorporate controls for the level of democracy. He claims that ‘overall, the choices of variables for these models does not change the difference in the relationship

* Binghamton University, SUNY (emails: davidc@binghamton.edu, mfilippov@binghamton.edu). Online appendices are available at <http://dx.doi.org/doi:10.1017/S000712341500071X>.

¹ Hill and Jones 2014.

of treaty ratification and respect for human rights'. We disagree. His findings show that whenever he estimates models including a control for the level of democracy, he finds no relationship between Treaty Ratification and the corrected scores (see his Figures 3–12, Models 3, 4 and 8). In other words, controlling for democracy, the relationships between treaty ratification and human rights are statistically *insignificant* for almost all measures of treaty ratification. When using the uncorrected scores, they are strongly significant in the *opposite*, negative direction. Thus with proper controls, the article actually shows the opposite of what the author claims. These crucial findings are never acknowledged or discussed in the article.

But why are the results obtained using the corrected scores of human rights as the dependent variable so sensitive to the controls for democracy? We conducted some additional analyses to investigate this question. It turns out that the apparent positive trend in the corrected scores that drives the results is largely due to an increase in the average level of democracy in the sample. Figure 2A (of our Appendix) groups the corrected scores by the level of democracy. Notice that for democracies (Polity IV \geq 6) the corrected scores recorded a decline since 1980. The performance of non-democracies was up and down with a little overall improvement. Thus if the average worldwide corrected scores are increasing over time, this is because the proportion of democracies in the sample goes up, almost doubling between 1965 and 2010 (from 30 to 58 per cent). To us, this is not clear evidence that there have been 'real improvements to the level of respect for human rights', as claimed by Fariss.

Our second concern is that the extrapolation of the data for the early years (1965–1980) drives the results reported in the article. We re-estimated all the models presented in the article for the period 1981–2010 (the range of the CIRI data) using the author's own data. The re-estimated regressions produced statistically insignificant differences between the coefficients obtained for the corrected and uncorrected scores (see Figures 3A–12A in our Appendix). Fariss' data extrapolation in the early years is based on very sporadic and eclectic bits of information about mass killings. Importantly, CIRI and PTS scores of human rights measure *domestic* repression, while Rummel, one of the important sources used in the extrapolation, recorded both international and domestic killings, making the records incomparable and many specific data points nonsensical.

For example, Fariss' corrected scores place the United States among the five worst violators in 1953 – above the Soviet Union, China, Albania and Czechoslovakia, but below North Korea. With one of the best scores, communist Mongolia, *de facto* occupied by the Soviet Union, consistently outperformed Switzerland. Sri Lanka outperformed the United Kingdom. Afghanistan trumped France. Ethiopia and East Germany scored better than the United States. We feel compelled to suggest that imputations so massive, and concentrated in a time period in which there were no data points for many countries, are not appropriate. This exercise comes close to data manufacture, and scores fabricated in this way should not be used in place of carefully collected and consistently coded real human rights data.

REFERENCES

- Christopher J. Fariss. 2016. 'The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance'. *British Journal of Political Science*, doi:10.1017/S000712341500054X.
- Hill, Daniel W. Jr., and Zachary M. Jones. 2014. An Empirical Evaluation of Explanations for State Repression. *American Political Science Review* 108:661–687.