IRT models for expert-coded panel data
Replication ReadMe

April 24, 2018

This dataset contains replication materials for "IRT models for expert coded panel data," which this document describes in detail. Note that the simulation analyses involve over 1,323 models (21 combinations of DIF and reliability × three bridging structures × three replications × seven models, plus several additional models in the appendix). Each of these models requires eight cores and 1-7 days to run using cluster computing resources at the Swedish National Infrastructure for Computing (OS: CentOS 6 and Ubuntu Xenial 16.04 LTS). As a result, while this replication data set includes the data and code necessary to reproduce the figures and tables in the paper—as well as posterior draws from the analyses of non-simulated data—it does not include the posterior draws from models run on simulated data, which are available from the authors upon request. Instead, we present the code necessary for replicating the simulation analyses in the manuscript, as well as the data that serve as the framework for our simulations.

Running replication code—excluding the run time for latent variable models—takes approximately 15 minutes (OS: Windows 7, 64 bit). Analyses require the following R libraries:

- coda
- gridExtra
- parallel
- reshape2
- rstan
- truncnorm
- tidyverse
- zoo

The replication data set includes six folders. The first folder, **Basic_Data**, contains the .RData files that serve as the basis for all analyses in the manuscript. The second folder, **Models**, contains the .R and .stan files for the models used in all analyses, as well as .RData files from latent variable models of actual (non-simulated) data. The third folder,

Actual_Data_Analyses, contains .R and .RData files to replicate the results of analyses of actual (non-simulated) data. The fourth folder, **Simulation_Analyses** contains .R files to replicate the simulated data sets, as well as .RData files containing the data used in the simulation analyses and descriptive statistics from the analyses in the paper. It also includes .R files for replicating figures in the manuscript related to the simulations. The remaining two folders include figures and tables.

To briefly note the files necessary to replicate specific graphics/tables:

- The file **actual_data_analysis.R** in **Actual_Data** contains code to produce Figures 1, 2, 3, E.1, E.2, and E.3; as well as Tables E.1, E.2 and E.3.
- The file **simulation_code.R** in **Simulation_Analysis** contains the code to produce Figures 7 and 8.
- The files **simulation_graphics_vdem**, **simulation_graphics_high_bridging** and **simulation_graphics_no_bridging** in **Simulation_Analysis** contain the code to produce Figures 4-6 and G.4, G.7 and G.10; Figures 9, 10, G.1, G.2, G.6, G.9 and G.12; and Figures G.3, G.5, G.8 and G.11; respectively.
- The file **ham_script.R** in **Simulation_Analysis** contains the code to produce Table C.1.
- The file **cauchy_script.R** in **Simulation_Analysis** contains the code to produce Tables H.1-4.

We discuss the contents of the folders in greater detail in turn.

1 Basic Data

This folder contains the .RData files that serve as the basis for all analyses in the manuscript. Specifically, it includes **v2clkill_reduced.RData**, the basis for all analyses in the paper; and **v2clkill_full.RData**, which contains the data we use to create country-date graphics.

The file **v2clkill_reduced.RData** contains coder-level data for the V–Dem v6.2 variable, "Freedom from Political Killings" (v2clkill). "_reduced" signifies that these data have been reduced from country-date to country-regimes, which we define as periods in which no experts change either their self-reported confidence in their coding or their actual codings. The object this file contains, "input.data" includes six objects itself:

wdata : Country-regime × coder data on regime trait (political killings); ordinal 1-5. Codes of -1 represent missing data.

N : Number of observations (country-regimes).

J : Number of coders.

K : Number of thresholds.

conf_mat : Country-regime × coder data on self-reported confidence; interval 0-1. Codes of −1 represent missing data.

cdata : Identifier of main-country coded for each coder (for use in hierarchical clustering of thresholds).

The file **v2clkill_full.RData** contains the non-reduced data for "Freedom from political killings." Note that this file does not contain cdata, since we do not run models with these data.

2 Models

We mainly model the country-regime latent trait—in both the actual (non-simulated) and simulated data— using six different latent variable models. This folder contains replication data for these models, as well as five additional models discussed in the appendix.

Each of the six main models has a set of three files: **.R**, the R file for running the analysis; **.stan**, the Stan model file, and **v2clkill.10000.x.RData**, posterior draws from x model run on the actual V–Dem data. We include the posterior draws from these models because the analyses take several days to run using cluster computing, with exact run times varying by model.

The names of the files designate the specific model:

rnone : IRT model that does not account for DIF or variation in coder reliability. Equation 4 in paper.

inone : IRT model that accounts for DIF with coder-specific hierarchical intercept parameter; does not account for variation in coder reliability. Equation 5 in paper.

none : IRT model that accounts for DIF with coder-specific hierarchical threshold parameters; does not account for variation in coder reliability. Equation 6 in paper. Note posterior distributions include a vestigial reliability parameter not included in estimation of latent trait.

rbeta : IRT model that does not account for DIF; accounts for variation in coder reliability with discrimination parameter. Equation 7 in paper.

ibeta : IRT model that accounts for DIF with coder-specific hierarchical intercept parameter; accounts for variation in coder reliability with discrimination parameter. Equation 8 in paper.

beta : IRT model that accounts for DIF with coder-specific hierarchical threshold parameter; accounts for variation in coder reliability with discrimination parameter. Equation 9 in paper.

aml : Bayesian Aldrich-McKelvey model. Appendix C; Equations C.1 and C.2.

We also include five additional models, which we discuss in the Appendices C (Hierarchical Aldrich-McKelvey models) and H (models with Cauchy distribution for universal thresholds). We do not include posterior draws from these models, since we did not run them on the actual V–Dem data.

rnone_cauchy : IRT model that does not account for DIF or variation in coder reliability. Universal thresholds use $Cauchy(0, 1)$ distribution.

inone_cauchy : IRT model that accounts for DIF with coder-specific hierarchical intercept parameter; does not account for variation in coder reliability. Universal thresholds use $Cauchy(0, 1)$ distribution.

rbeta_cauchy : IRT model that does not account for DIF; accounts for variation in coder reliability with discrimination parameter. Universal thresholds use $Cauchy(0, 1)$ distribution.

ibeta_cauchy : IRT model that accounts for DIF with coder-specific hierarchical intercept parameter; accounts for variation in coder reliability with discrimination parameter. Universal thresholds use $Cauchy(0, 1)$ distribution.

ham : Hierarchical Aldrich-McKelvey model. Equations C.3, C.4 and C.5.

Note that the R files use **v2clkill_reduced.RData** as their input data. To analyze simulated data, change the input data name to that which is relevant.

3 Actual Data

The file **actual_data_analysis.R** contains code to replicate Figures 1, 2, 3, E.1, E.2, and E.3; as well as Tables E.1, E.2 and E.3. This file takes approximately two minutes to run, and requires both files in the folder **Basic_Data** and the posterior draws in **Models**. **dataset_actual_data.RData** contains a dataset of point estimates (posterior medians) and credible regions (lower and upper 95 percent HPD interval bounds) from the latent trait from different models, named according to the earlier schema. The dataset also contains the normalized average of country-year scores (wca) and upper and lower bounds for 95 percent confidence intervals; as well as an indicator (reduced) for whether or not a country-year is included in the reduced data (1 =reduced).

4 Simulation Analysis

This folder contains the code necessary to create the simulated datasets we use in our analyses, as well as the datasets themselves. It also contains a dataset of statistics based on the analyses we ran, as well as code to create these statistics. Finally, it includes files to prepare the graphics in the manuscript and appendix.

The file **simulation_code.R** provides the code for replicating our simulation analyses, and takes approximately four minutes to run. The code is divided into four sections. The first section provides the code to simulate various forms of DIF and reliability. This section

requires the file **v2clkill_reduced.RData** from the **Basic_Data** folder. The second section uses this simulation framework to produce simulated data using the V–Dem data structure, with different combinations of DIF and reliability. Note that code for all combinations of reliability and DIF save one (no variation in reliability and no DIF) are commented out; to produce simulated data with different combinations of DIF and reliability run the relevant section (and comment out the running version). The third section provides code for producing a dataset structure of high bridging and the fourth code for a dataset structure of no bridging; both sections conclude with the bridging analysis code used to create Figures 7 and 8. To create data with different bridging patterns and combinations of reliability and DIF simply run the relevant code in the second section after that for the third or fourth. Finally, to replicate results for a given model and combination of simulated reliability/DIF and bridging structure, first create the relevant simulated data and run the desired model in the folder **Models** on these data. Note that this will require changing the name of the input data in the relevant model's R file.

The nine files **v2clkill_simdata_x_y.RData** are the framework files we used in our simulation analyses, where x represents the data structure (blank refers to the V–Dem data structure, *hbridging* high bridging, and *nbridging* no bridging) and y an arbitrary number representing the same patterns of reliability and DIF (all simulations were conducted thrice). **simulation_dataset_stats_script.R** provides the code we use to analyze all model output. In lieu of providing posterior draws from the 1,323 latent variable models of different combinations of simulated data, we provide **dataset_simulation.RData**, which contains relevant statistics across all simulations. Descriptions of column names are as follows:

Model	Model used to aggregate data. "Mean" is the normalized country-regime mean of the simulated data, "AM" is the Bayesian Aldrich–McKelvey algorithm, "No DIF" is an IRT model without DIF parameterized, "Intercept DIF" is an IRT model with intercept DIF parameterization, and "Threshold DIF" is an IRT model threshold DIF parameterization.
Data	Simulation data structure. "V–Dem" is the V–Dem data structure; "HBridging" is high-bridging and "NBridging" no bridging.
Simulation	Arbitrary simulation replication number, 1-3.
Simulated_DIF	Form and magnitude of simulated DIF.
mulated_Reliability	Form and magnitude of simulated DIF.
Model_Reliability	Whether model fixed expert reliability or allowed it to vary.
Converged	Indicator if latent trait parameters converged: 95 percent or more parameters had $\hat{r}$ scores of 1.1 or lower according to Gelman-Rubin diagnostic.
MSE	Mean-squared error of point estimates (posterior median) from true values.
Pearson	Pearson correlation statistic of point estimates with true values.
Kendall	Kendall correlation statistic of point estimates with true values.

HPD Proportion of true values contained in 95 percent HPD interval.

Iterations Number of model iterations.

Note that any additional replication results will differ from those in **dataset_simulation.RData** based on version of Stan and R, as well as seeds set; a successful replication would yield results that do do not substantially differ from those which we present. Also note that this file does not contain the data from the additional models discussed in the appendix (Hierarchical A–M and models with Cauchy priors on the universal thresholds). Instead, data from the Cauchy models are contained in **cauchy_data.RData** and Hierarchical A–M data in **ham_data.RData**. To produce Table C.1, use the file **ham_script.RData**; to produce Tables H.1, H.2, H.3, and H.4 use the file **cauchy_script.RData**. Both files take less than a minute to run. Note that the tables that **cauchy_script.RData** produces order statistics by simulation version; in the manuscript data from models with uniform priors on the universal thresholds are ordered median estimate (lower estimate, upper estimate).

The **simulation_graphics_x.R** files provide code to produce the graphics in the paper, with x corresponding to different data structures. Each file takes approximately a minute to run. **_vdem** contains the code to produce Figures 4-6 and G.4, G.7 and G.10; **_high_bridging** Figures 9, 10, G.1, G.2, G.6, G.9 and G.12; **_no_bridging** Figures G.3, G.5, G.8 and G.11.