The Determinants of Uncertainty in International Relations

Data Codebook

Jeffrey M. Kaplow; William & Mary; jkaplow@wm.edu
Erik Gartzke; University of California, San Diego; egartzke@ucsd.edu

This data codebook describes the replication package for "The Determinants of Uncertainty in International Relations." Additional detail about the coding of variables and data sources is available in the article.

Contents of replication package

A Stata `.do` file to replicate the statistical results from the article is enclosed (`DetUnc_Replication.do`), along with an `.R` file (`DetUnc_Figure_Replication.R`) to replicate the figures. Analyses were conducted using StataSE version 4.2. Figures were created with *R* version 3.5.2. The replication package includes five data files:

`retrodat.dta`	The full retrospective uncertainty dataset
`overdat.dta`	Retrospective uncertainty data limited to over-estimates
`underdat.dta`	Retrospective uncertainty data limited to under-estimates
`coderdat.dta`	The full coder-assessed uncertainty dataset
`DetUnc_Figure_Replication.rdata`	Data used for the figures in the article

Contents of Stata data files

The Stata `.dta` files in the replication package contain the variables described below. For the sources used to derive each variable, please see the main article. Our empirical analysis employs a within-between random effects approach. "Between" variables are denoted with a `_bw` suffix; "within" measures have a `_w` suffix. Please see the main article for more details.

`ccode`	Correlates of War country code for the state in the dyad
`year`	Observation year
`retro_frc`	Retrospective measure of armed forces personnel

retro_frc_lag	Retrospective measure of armed forces personnel (lagged one year)
first_pers	Earliest estimate of the number of personnel in the states' armed forces in that year
rgdppc	Real GDP per capita
hasnuke	Dichotomous variable set to 1 if a state has a nuclear weapon in that year
tdriv	Dichotomous variable set to 1 if a state has an enduring rivalry in that year
ally	Dichotomous variable set to 1 if a state has a formal defense pact in that year
midslast5dum	Dichotomous variable set to 1 if a state has participated in a militarized interstate dispute in the previous 5 years
polity2	Measure of state regime type
coder_frc_any	Measure of coder-assessed uncertainty (estimate marked as uncertain or omitted)
coder_frc_any_years	Count of years since coders marked an estimate as particularly uncertain; squared (coder_frc_any_years2) and cubed (coder_frc_any_years3) counts are also used
wmeatvol	Year of the WMEAT volume used to construct the retrospective measure (the more recent of the two volumes)
polity_ryear	Measure of state regime type at the time of reassessment
NATO	Dichotomous variable set to 1 if a state is a NATO member in that year

Contents of R data file

The R DetUnc_Figure_Replication.rdata file in the replication package contains the data structures described below.

alldat	A data frame containing country-year retrospective and coder assessed measures of uncertainty (used in Figure 1); includes the following variables:

`ccode`	Correlates of War country code for the state
`year`	Observation year
`retro`	Retrospective measure of uncertainty; negative values reflect underestimates, positive values reflect overestimates
`coder`	Coder-assessed measure of uncertainty

`subeffects`   A list containing the data structures necessary for plotting substantive effects (used in Figures 2 and 3), derived from the analysis in Stata provided in the `DetUnc_Replication.do` file; includes the following data frames:

`first_pers`	A data frame containing substantive effects on retrospective uncertainty for the earliest estimate of the number of personnel in the states' armed forces in a particular year (used in Figure 2)
`rgdppc`	A data frame containing substantive effects on retrospective uncertainty for real GDP per capita (used in Figure 2)
`polity2`	A data frame containing substantive effects on retrospective uncertainty for state regime type (used in Figure 2)
`between`	A data frame containing substantive between-country effects on coder-assessed uncertainty (used in Figure 3)
`within`	A data frame containing substantive within-country effects on coder-assessed uncertainty (used in Figure 3)

The `first_pers`, `rgdppc`, and `polity2` data frames contain the following variables (see the discussion of Figure 2 in the article for details):

`pp/ppo/ppu`	The predicted level of retrospective uncertainty associated with a particular value for this variable (given in `x`, below) among all estimates (`pp`), over-estimates (`ppo`), and underestimates (`ppu`)
`cil/cilo/cilu`	The lower boundary of the 95% confidence interval associated with a particular value for this variable (given in `x`, below) among all estimates (`cil`), over-estimates (`cilo`), and underestimates (`cilu`)
`cih/ciho/cihu`	The upper boundary of the 95% confidence interval associated with a particular value for this variable

(given in `x`, below) among all estimates (`cih`), over-estimates (`ciho`), and underestimates (`cihu`)

x   The particular value of the variable at which substantive effects are estimated

The `between` and `within` data frames contain the following variables, with each row representing a particular factor in the data analysis (see the discussion of Figure 3 in the article for details):

fd   The change in the predicted probability of coder-assessed uncertainty when a factor is shifted from two standard deviations below its mean to two standard deviations above its mean, holding all other variables at the mean among those observations with a greater than 5 percent chance of uncertainty

cil   The lower boundary of the 95% confidence interval

cih   The upper boundary of the 95% confidence interval