Data Codebook for "Forecasting the Onset of Genocide and Politicide: Annual Out-of-Sample Forecasts on a Global Dataset, 1988-2003"

Goldsmith, Butcher, Semenovich, and Sowmya

JPR

---

variable name in dataset – "Variable name in article"; Variable description; Data source.[1]

ccode – [not mentioned in article]; Numerical country identifier; Correlates of War (COW) dataset (Ghosn and Bennett 2007).

year – Current year for outcome variables; Ranges from 1974 through 2003, year "*t*", while the corresponding "*t-1*" values for predictor variables are 1973 through 2002; PITF dataset (Marshall, Gurr, and Harff 2010).

ef – "EF_{t-1}"; Ethnic fractionalization; Data are from Fearon (2003).

instability – "$Instability_{t-1}$"; Political instability in year *t-1*; Political Instability Task Force (PITF) dataset (Marshall, Gurr, and Harff 2010).

finstability – "$Instability_{t}$"; Political instability in year *t*; PITF (Marshall, Gurr, and Harff 2010).

lntotalpop – "$ln Population_{t-1}$"; Total population, natural logarithm; COW dataset (Ghosn and Bennett 2007). .

mena – "MENA"; Middle East and North African states (MENA); based on COW (Ghosn and Bennett 2007).

csasia – "CSASIA"; Central and South Asian states (CSASIA); based on COW (Ghosn and Bennett 2007).

nac – "$NeighborConflict_{t-1}$"; Number of bordering states with any type (societal or interstate) internal conflict; Major Episodes of Political Violence dataset (Marshall 2010).

abbrev – [not mentioned in article]; Country abbreviation; COW (Ghosn and Bennett 2007).

country – [not mentioned in article]; Country name; COW (Ghosn and Bennett 2007).

---

[1] Abbreviations and bibliographic references are as in main article text.

fgenpold – [not mentioned in article]; Genocide / Politicide dummy variable year t+1, including ongoing cases; PITF (Marshall, Gurr, and Harff 2010).

AssassinDum – "$Assassin_{t}$"; Successful or attempted political assassination, dummy variable for any instance(s) in a given year; Banks (2011).

stabyrs – "$StabYrs_{t-1}$"; Time since last instability year, begin from 1948, calculated with BTSCS software (available here: http://www-personal.umich.edu/~wmacmill/stata/btscs/btscs-a-binary-time.pdf); Based on PITF data (Marshall, Gurr, and Harff 2010).

_prefailstab - "$PrevInstability_{t-1}$"; Number of previous failures (instability years), calculated with BTSCS software (see URL above); Based on PITF data (Marshall, Gurr, and Harff 2010).

stabyrs2 – "$StabYrs^{2}_{t-1}$"; squared stabyrs; Based on PITF data (Marshall, Gurr, and Harff 2010).

polity2sq_inv – "$MixedRegime_{t-1}$"; The polity regime-type index, ranging from -10 through +10, squared and subtracted from 100 such that it takes higher values for regimes towards the middle of the index; Polity IV dataset (Marshall, Jaggers and Gurr 2010).

NUMIGO – "$NUMIGO_{t-1}$"; Number of International Governmental Oorganization memberships; COW IGO dataset (Pevehouse, Nordstrom, and Warnke 2004).

regchange1to3 – "$PolChange3_{t-1}$"; Regime change over last three years, calculated as the sum of first differences [($Polity_{t-1}$ - $Polity_{t-2}$) + ($Polity_{t-2}$ - $Polity_{t-3}$) + ($Polity_{t-3}$ - $Polity_{t-4}$)]; Polity IV dataset (Marshall, Jaggers and Gurr 2010).

lIMR – "IMR_{t-2}"; Lagged Infant Mortality Rate, *t-2*; Data are from Abourharb and Kimball (2007).

lAssasinDum – "$Assassin_{t-2}$"; Lagged Political assassination *t-2*, Successful or attempted assassination; Banks (2011).

SLDdum – "SLD_{t-1}"; State-led Discrimination, either economic or political, against a particular group, coded as a dummy variable with state-years not included in the MAR dataset = 0; Minorities at Risk dataset (2009).

genpold – [not mentioned in article]; Genocide / Politicide dummy variable year t, including ongoing cases; PITF (Marshall, Gurr, and Harff 2010).

statename – [not mentioned in article]; State (country) name, shortened or adjusted from "country" variable; authors.

fgenpolonsetd – "$GenocideOnset_{t}$"; Genocide / Politicide onset, ongoing years set to missing; PITF data (Marshall, Gurr, and Harff 2010).

_prefailgp – "PrevGen$_{t-1}$"; Number of previous failures (genocide/politicide onsets) since 1948, calculated with BTSCS software (see URL above); Based on PITF data (Marshall, Gurr, and Harff 2010).

nogpyrs2 – "NoGenYrs$^2_{t-1}$"; Squared value of time since last genocide/politicide onset, begin 1948, calculated with BTSCS software (see URL above); Based on PITF PITF data (Marshall, Gurr, and Harff 2010).

nogpyrs3 – "NoGenYrs$^3_{t-1}$"; Cubed value of time since last genocide/politicide onset, begin 1948, calculated with BTSCS software (see URL above) from PITF data (Marshall, Gurr, and Harff 2010).

RulingEliteEthnicity – "EthnicPowerRel$_{t-1}$"; Values: 0 ethnicity not relevant, 1 ruling elite in the majority, 2 ruling elite in the minority; Ethnic Power Relations dataset (Cederman, Min, and Wimmer 2009) .

fullaut – "FullAutt-1"; Fully authoritarian regime dummy variable, Polity values of -7 or less are coded as fully authoritarian; Polity IV dataset (Marshall, Jaggers and Gurr 2010).

partaut – "PartAut$_{t-1}$"; Partially authoritarian regime dummy variable, Polity values from -6 through 1 are coded as partially authoritarian; Polity IV dataset (Marshall, Jaggers and Gurr 2010).

partdem – "PartDem$_{t-1}$"; Partially democratic regime dummy variable, Polity values from 2 through 6 are coded as partially democratic; Polity IV dataset (Marshall, Jaggers and Gurr 2010).

humdefburCC – "HDB$_{t-1}$"; Human defense burden, calculated as per Colaresi and Carey (2008), total military personnel divided by natural log of total population; COW dataset (Ghosn and Bennett 2007).

exuncHDB_cc – "Uncon×HDB$_{t-1}$"; as per Colaresi and Carey (2008), unconstrained executive interacted with (multiplied by) human defense burden, we interact a dummy variable indicating an unconstrained executive (scoring 1 or 2 on the exconst component of Polity) with the number of military personnel divided by the logged population. Polity IV and COW dataset (Ghosn and Bennett, 2007).

FDexuncHDB_cc – "Uncon×fdHDB$_{t-1}$"; first difference of "exuncHDB_cc ", (*t-1*) – (*t-2*); Polity IV and COW dataset (Ghosn and Bennett, 2007).

NELDAperiod – "ElectionPeriod$_{t-1}$"; Election Period, there is an election recorded in the NELDA data for year *t-2*, *t-1*, or *t*; NELDA dataset (Hyde and Marinov 2012).

efNELDAperiod – "EF_{t-1}×$ElectionPeriod_{t-1}$"; Ethnic fractionalization interacted with (multiplied by) Election period; data sources EF (Fearon 2003) and NELDA (Hyde and Marinov 2012).

fNELDAperiod – "$ElectionPeriod_t$"; Forward-lagged Election Period, i.e., there is an election recorded in the NELDA data for year *t-1*, *t*, or *t+1*; NELDA dataset (Hyde and Marinov, 2012).

impute01 – [not mentioned in article]; Variable marking observations (country-years) with imputed data, these observations are characterized in the article text; authors.

yeartminus1 - [not mentioned in article]; Year of predictors, i.e., all variable names above not beginning with the letter 'f' correspond to this year; variable names beginning with the letter 'l' (but not 'ln') are for year *t-2*;

pr_instab – "$PrInst_t$" ; this variable is not included in the dataset, but is generated from the stage 1 model using the Stata command "predict", which produces the estimated probability of instability in year *t*; authors.

Summary statistics for the data file:

Variable	Obs	Mean	Std. Dev.	Min	Max
ccode	5949	467.8139	261.8754	2	990
year	5949	1988.457	8.655013	1974	2003
ef	4276	.4779719	.265302	.001999	1
instability	5776	.1497576	.3568646	0	1
finstability	5766	.1498439	.3569493	0	1
lntotalpop	5722	15.29908	2.072404	9.305651	20.98563
mena	5776	.1101108	.3130549	0	1
csasia	5776	.0727147	.2596901	0	1
nac	4380	.9954338	1.22767	0	7
abbrev	0				
country	0				
fgenpold	5764	.0324427	.1771882	0	1
AssassinDum	4296	.1117318	.3150729	0	1
stabyrs	5776	21.42452	14.86753	0	54
_prefailstab	5776	3.569945	7.18589	0	54

stabyrs2	5776	680.0149	683.86	0	2916
polity2sq_~v	4404	43.20005	31.17046	0	100
NUMIGO	5059	50.02471	21.52846	1	129
regchange1~3	4287	.4795895	3.22496	-16	18
lIMR	4914	58.77911	48.08093	2.3	284.78
lAssassinDum	4265	.1099648	.3128826	0	1
SLDdum	5776	.2657548	.4417726	0	1
genpold	5776	.0339335	.1810738	0	1
statename	0				
fgenpolons~d	5778	.0044998	.0669354	0	1
_prefailgp	5949	.1563288	.4537474	0	3
nogpyrs2	5949	1025.232	711.5449	0	2916
nogpyrs3	5949	38698.1	37274.95	0	157464
RulingElit~y	4074	.9280805	.5382332	0	2
fullaut	4404	.3626249	.4808124	0	1
partaut	4404	.1943688	.3957589	0	1
partdem	4404	.1049046	.3064653	0	1
humdefburCC	5038	8413.623	22142.68	0	229322.2
exuncHDB_cc	4379	2088.456	8335.537	0	208556.6
FDexuncHDB~c	4345	-59.41858	3915.518	-208334.5	50754.61
NELDAperiod	5949	.4654564	.4988472	0	1
efNELDAper~d	4276	.2823244	.3013568	0	1
fNELDAperiod	5946	.4724184	.4992807	0	1
impute01	72	1	0	1	1
yeartminus1	5949	1987.457	8.655013	1973	2002