Appendix to "Institutional Inconsistency and Political Instability: The Duration of Polities"

Author

December 19, 2005

1 Introduction

This codebook documents the data used in Gates, Scott G., Håvard Hegre, Mark P. Jones, and Håvard Strand, "Institutional Inconsistency and Political Instability: Persistence and Change in Political Systems Revisited, 1800-2000". First, we discuss the two different data structures used, and how they are calculated. We then go on to present the different variables of the dataset in detail.

2 MIRPS Data Structure

The unit of analysis in the MIRPS dataset is a polity, defined by Gurr (1974: 1483) as "the basic political arrangement by which national political communities govern their affairs". It is the duration of this arrangement that in effect is our dependent variable. Our operationalization is based on three dimensions:

1. Decision constraints: XCONST variable from the Polity data project. Our data is mainly based on the Polity IIId (McLaughlin et al., 2003) dataset, with additional data from Kristian S. Gleditsch' (2003) modified dataset (version 1.0), and information on precise dates gathered by the Authors and research assistants.

2. Executive recruitment: XRREC, XROPEN, and XRCOMP. These variables are gathered from the same sources as the previous dimension.

3. Participation: We use Vanhanen's (2000) Polyarchy data as described in the paper.

All data sources are supplied with dates of varying precision. Based on these dates, we define the duration of a polity from a start date to an end date. Since there always is some form of change within the definition of polity, we define these dates based on a rubber band algorithm. When a new polity is defined, we record the current values at that day for all three dimensions. We then define an area within which the values on these dimensions can vary. When one or more dimensions go outside this area, we define a new polity, with new values and variance area. Since we operate with highly time-varying variables on the right-hand side, we divide each polity up in yearly observations and censor each time period which does not contain a regime change. All lines in the dataset are located in time by the variables STARTND and ENDND, which measures the number of days since 1.1.1800 to the first and last day of the observation, respectively. It is therefore necessary to define a censor variable, STATUS, which takes 1 where a regime change happens and 0 for all censored time periods. In order to analyze these separated units as one statistical unit, we construct a variable STSETPOLID containing uniqe IDs for all polities. For the first observation within a polity, we define a variable ENTRYDATE, which holds the date of birth for that polity. The dataset is finally prepared for analysis in Stata by the stset command:

```
stset endnd, id(stsetpolid) failure(status==1) origin(time entrydate) scale(365.25)
```

3 Definition of Polity Change (repeated from section 3.2)

We define a polity change as any change in indicators that results in either: (1) a movement from one category to another in the Executive dimension (i.e., between ascription/designation, dual ascriptive/elective, and elective), (2) a change of at least two units in the Executive Constraints dimension, or (3) a 100% increase or 50% decrease in the Participation dimension (in the log-transformed variable, this is a change of 0.69 in either direction from the original level). Doubling the number of citizens with voting rights qualifies as a minimum change along the Participation dimension. The creation or dissolution of states is also defined as a polity change.

4 Przeworski et al. data structure

The analysis is based on two different data structures. In Table 3, Model 3 and 4, we analyze an alternative data structure, based on Przeworski et al. (2000) (hereafter PACL). The basic unit in this data structure is the country-year, and the dependent variable is whether a regime change happened during that year or not. The country-year observation is very common in quantitative analysis of time series (see Beck, Katz and Tucker, 199x), but is has certain short-comings. The most relevant difference with our research design is that in is structure each observation can only contain one value on the regime variable. This becomes a problem in years where a regime changes.

The PACL dataset reports the regime statistics for 31 December of every year. If we observe a regime change in a given year, the PACL dataset will not report the regime characteristic of the regime that failed but of the new regime which was created. This is the opposite of what we want and what we do in our own data structure. We must therefore lag the PACL regime variable one year. Since our SIP scalar can vary within the unit of observation present, we use the unit's SIP value as of January 1 of the relevant year. An example of this is Chile, 1973.

Table 1: Chile, 1973

Country	Status	SIP	YEAR	dd_reg	dd_lagreg	dd_reg_id	dd_lagreg_id
Chile	0	0.8138	1972	0	0	15501	15501
Chile	1	0.8138	1973	1	0	15502	15501
Chile	0	0	1974	1	1	15502	15502

The democracy of Chile failed in 1973, and the PACL regime variable (dd_reg) does therefore indicate that Chile is a dictatorship at the end of 1973. However, our interest lies with the regime that failed, not the regime that was created, so we lag the regime characteristics (dd_lagreg) and the regime id variable (dd_lagreg_id). The Status variable indicates that there was a regime change in Chile in 1973. As with the MIRPS data structure, we divide each political regime into several observations, and add an identifying variable (dd_reg_id) that allows us to analyze. As shown above, we lag this variable as well. This data structure is prepared for analysis using a similar command as the previous data structure.

```
stset endnd, id(dd_lagreg_id) failure(status==1) origin(time entrydate) scale(365.25)
```

2

5   List of variables

ssno A number identifying which state the polity is existing within. The numbering system used here is based on Kristian Gleditsch's updating of item[Polity IIId, which in turn is based on the Gleditsch and Ward (1999) System Membership definition.

polid Since each country can have several polities during it's history, we separate different polities from each other through different Polity IDs

startd The start date as a DATE type variable

endd The end date as a DATE type variable. This, and the previous variable, is only included in order to ease the readability of the dataset.

startnd The start date represented as the number of days since 1 January 1800

endnd The end date represented as the number of days since 1 January 1800

xconst The 'Executive Constraints' dimension (Polity IV codebook, p.21)

xrec The 'Executive Recruitment' dimension. This variable is constructed out of three Polity IV indicators: XRCOMP, XRREC, and XROPEN.

part The 'Participation' dimension. This variable is calculated from Tatu Vanhanen's (2000) Polyarchy dataset. Primarily, we define this dimension parallel to Vanhanen's participation indicator, which measures the fraction of the population which participated in an election. However, since we theoretically are interested in capturing the extent to which the capacity of changing the composition of the government is distributed in a society, we have taken the effective competition in the election into account, so that the participation score is multiplied with the fraction [Competition/30

status This variable defines whether the polity ended on the end date or not. Our data end at 31 December 2000, and all polities in existence at this day are ended. However, this should not be analyzed as the end of a polity, since it is just the end of us observing the polity. This is often referred to as 'censoring' in the survival analysis literature. For analytical purposes, it is practical to add a number of censored observations, which effectively increases the number of control cases, and which further improves the strength of the analysis. All these additional control cases will be recorded with status equal to 0, similar to those observations that end on 3 December 2000. For the analysis in "Institutional Inconsistency and Political Instability: Polity Duration, 1800-2000", we censor every observation at the end of every year, which consequently means that the number 0's vastly outnumbers the 1's.

duration The duration of that polity, measured in the number of days.

xconsdic Dichotomized Xconst-variable $(1 : \{1 - 4\} ; 2 : \{5 - 7\})$ (not longer used)

particip A coding of participation into a ordinal variable (not longer used)

xrreg Polity IV variable xrreg

xrcomp Polity IV variable xrcomp

xropen Polity IV variable xropen

sip2 Sip index.

sip2avgnabo The average of the difference between the country in question and all neighbouring countries, individually, following the expression The neighborhood is defined as all countries either contiguous or separated by less than 25 nautical miles of sea to the country in question.

sip2varnabo The standard deviation of the neighbourhood's SIP2 scores.

reg

sip2status (not in use)

sip2ysc Calender years since the previous change in the SIP2 score

sip2_previous The SIP2 score of the previous polity

stsetpolid Variables for specifying Stata analysis

stsetorig —————- " —————

6   Control variables

GDP per capita data were drawn from three sources. We use World Bank data for the period 1960 to 1998 (World Bank, 2000a; 2000b), Penn World Tables, v5.6 (Summers and Heston, 1991) for 1950 to 1959, and Maddison's (1995) Monitoring the World Economy 1820-1992 for the years 1900-1949. The three datasets refer to different baseline years for calculating constant dollar figures, and are based on different methods of measurement. To counter these differences, we calculate the average ratio in the three first overlapping years per country for both overlaps, and use this ratio to adjust the numbers. To reduce endogeneity bias, we lag the variable. We use the average ln(GDP per capita) for the five years preceding the end-date of each time segment

7   Calculation of the demindex and SIP variables

We provide two different democracy indexes, which almost correlate perfectly, but are calculated in two different ways. The 'demindex' variable is the length of the vector D= [(0,0,0),(xconst, xrec, part)]. This length is determined using

$$|D| = \sqrt{(xconst^2 + xrec^2 + part^2)} \tag{1}$$

The sip score is defined as the value of the coordinate x of the point (x,y,z) on the diagonal vector ([(0,0,0),(1,1,1)]), where the following is true:

$$(x, y, z) \cdot (x - xconst, y - xrec, z - part) = 0 \tag{2}$$

Since this point is on the diagonal, x, y, and z must all be equal. We can therefore write

$$(x, x, x) \cdot (x - xconst, x - xrec, x - part) = 0 \tag{3}$$

$$x \cdot (x - xconst) + x \cdot (x - xrec) + x \cdot (x - part) = 0 \tag{4}$$

$$x^2 - x \cdot xconst + x^2 - x \cdot xrec + x^2 - x \cdot part = 0 \tag{5}$$

$$3x^2 - x \cdot xconst - x \cdot xrec - x \cdot part = 0 \tag{6}$$

$$3x^2 = x \cdot xconst + x \cdot xrec + x \cdot part \tag{7}$$

$$3x = xconst + xrec + part \tag{8}$$

$$x = SIP = \frac{xconst + xrec + part}{3} \tag{9}$$

The sip index is normalized to the range [0,1]

8 Polity IV categories

We base two of our indicators, Executive Contraints and Executive Recruitment, on the Polity IV dataset. The Executive Contrainsts dimension is based on the XCONST variable alone, and does therefore not demand any further focus. The Executive Recruitment dimension (XREC in our dataset) is based on three polity variables, XRCOMP, XRREG, and XROPEN. Table 1 documents the coding of the XREC dimension. Table 2 lists the distribution of the XCONST variable. The number of polities in these tables refers to the Polity IV definition of a polity, and is therefore substantially different from a similar count based on the replication data documented herein. Partially this is due to the fact that a small regime change in the polity database does not result in a regime change by our definition. On the other hand, a regime change in the replication data can be triggered by a change in the participation measure, which of course is unnoticed in the Polity database. Since these numbers are based on incrementally adapted versions of the Polity dataset, we do not expect these tables to be perfectly reproduced using data from the Polity project webpage. See section 3 for the definition of a polity change.

Table 2: Definition of Authority Dimension Executive Recruitment

Cat. no.	Polity II label	Brief definition	XRREG	XRCOMP	XROPEN	Number of polities
(NA)	Missing data-/transition	(excluded from analysis)	-66, -77, -88, -99	-66, -77, -88, -99	-66, -77, -88, -99	242
1	Caesaristic	Self-selection by seizure of power	1 Unregulated	0 Unregulated	0 Unregulated	150
		Shifts to/from such polities	2 Designational / Transitional	0 Unregulated	0 Unregulated	48
2	Designation	Informal	2 Designational / Transitional	1 Selection	4 Open	340
		Intra-elite competition	1 Unregulated	1 Selection	4 Open	3
	Ascription	Succession by birthright	3 Regulated	1 Selection	1 Closed	95
	Dual Exec: Ascription + Designation	Ascriptive and designated rule coexist	3 Regulated	1 Selection	2 Dual executive - Designation	95
3	Dual Executive: Ascription + Election	Ascriptive and elected rulers coexist	3 Regulated	1 Selection	3 Dual executive-Election	3
			2 Designational-/ Transitional	2 Dual/ Transitional	4 Open	105
			3 Regulated	2 Dual/ Transitional	3 Dual executive-Election	10
			3 Regulated	2 Dual/ Transitional	4 Open	5
4	Election	Formal competition among popularly supported candidates	3 Regulated	3 Election	4 Open	302

Table 3: Definition of Authority Dimension Executive Constraints

Category number	Brief Definition (XCONST)	Number of polities
(missing)	-66, -77, -88, -99	242
1	Unlimited executive authority	321
2	Intermediate category	97
3	Slight to moderate limitations on executive authority	337
4	Intermediate category	33
5	Substantial limitations on executive authority	113
6	Intermediate category	46
7	Executive parity or subordination of executive authority	209