Thomas Plumper and Eric Neumayer

The level of democracy during interregnum periods: recoding the polity2 score

Article (Accepted version) (Refereed)

Original citation:
Plumper, Thomas and Neumayer, Eric (2010) The level of democracy during interregnum periods: recoding the polity2 score. Political analysis, 18 (2). pp. 206-226. ISSN 1047-1987
DOI: 10.1093/pan/mpp039

The Level of Democracy during Interregnum Periods:

Recoding the *polity2* Score

Published in:

Political Analysis, 18 (2), 2010, pp. 206-226

Thomas Plümper[a] *and Eric Neumayer*[b]

[a] Department of Government, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK, and Centre for the Study of Civil War, International Peace Research Institute Oslo (PRIO), tpluem@essex.ac.uk

[b] Department of Geography and Environment, London School of Economics and Political Science, London WC2A 2AE, UK, and Centre for the Study of Civil War, International Peace Research Institute Oslo (PRIO), e.neumayer@lse.ac.uk

Corresponding author: Thomas Plümper. Equal authorship. We thank three anonymous referees, the editors as well as Kristian S. Gleditsch, Nils Petter Gleditsch, Håvard Strand and Michael D. Ward for helpful comments. The data for the replication analyses as well as our alternative polity variables will be made available for download from http://personal.lse.ac.uk/neumayer or http://www.polsci.org/pluemper/.

The Level of Democracy during Interregnum Periods:

Recoding the *polity2* Score

Abstract

The *polity2* variable from the Polity IV project is the most popular measure of a country's political regime. This article contends that the coding rules employed to create a *polity2* score during years of so-called interregnum and affected transitions produce a measure of democracy that lacks face validity. Using both single and multiple imputation methods, we construct and evaluate several variables that offer alternative measures to *polity2* during such periods. We recommend that scholars using *polity2* test whether their results are robust to using our alternatives and using multiple imputation techniques instead. Where robustness cannot be established, scholars need to theoretically justify the choice of either *polity2* or one of the alternatives.

1. Introduction

Of the data sets available to scholars researching questions relating to democracy, the Polity data set (Jaggers and Gurr 1995; Marshall and Jaggers 2002) is by far the most popular one. Social scientists in general, and political scientists in particular, are much more likely to trust Polity as their main source of information and use, if at all, one of its competitors as an additional robustness check.[1] There are many good reasons for Polity's apparent dominance. It offers the broadest coverage of all democracy indicators, including 187 countries from either 1800 or the year of independence up to 2008, currently the latest year available. It also relies on a fairly comprehensive definition of democracy, which includes electoral rules and various measures of the openness of political institutions, and provides detailed information on aspects of institutionalized democracy and autocracy in a country at any point of time.

Notwithstanding this wealth of information, the *polity* variable, a composite score of the various variables included in the Polity data set, is by far the most commonly used information. Indeed, most researchers have a theoretical interest in the causes and consequences of a country's level of democracy. No surprise, then, that to many scholars the data set became even more attractive when – with the move from the Polity III (Jaggers and Gurr 1995) to the Polity IV (Marshall and Jaggers 2002) version – a new variant of the polity score, called *polity2*, was added. While both variables are in principle computed by subtracting Polity's institutionalized autocracy score (*autoc*) from its institutionalized democracy score (*democ*) to generate an aggregate democracy variable that runs from -10 to 10, the *polity2* variable has one main advantage over *polity*: it seemingly provides a

[1] See Munck and Verkuilen (2002) and Strand (2007) for a review and evaluation of available data sets. For robustness test with different democracy scores see Casper and Tufis (2003).

democracy score for periods of so-called ‘interregnum’ and ‘transition’, whereas *polity* coded these with placeholder values as -77 and -88, respectively, and therefore as essentially missing. In developing the *polity2* variable, the principal investigators of Polity IV decided on simple rules for coding the political regime in events of ‘interregnum’ and ‘transition’. The explicitly stated aim of the *polity2* coding is “to facilitate the use of the POLITY regime measure in time-series analyses” (Marshall and Jaggers 2007: 8).

This article contends that the *polity2* coding rules for the creation of what seems to be a measure of the level of democracy during periods of ‘interregnum’ and ‘affected transitions’[2] are problematic and produce democracy scores for some affected country years that lack face validity. They also have the potential to invalidate causal inferences from quantitative analyses using the *polity2* variable. This becomes more likely if researchers use *polity2* as an explanatory variable in a quantitative analysis in which the dependent variable is civil war or state collapse.

Our paper makes two contributions. First, we alert potential users of *polity2* to the necessity of paying careful attention to the problematic treatment of countries during periods of interregnum and affected transitions. Though we do not claim to have a unique better solution for this imputation problem, we contend that the several alternative imputation strategies we suggest result in alternative variables to *polity2* that have higher face validity and treat the uncertainty associated with the imputed values more appropriately.

Our second contribution is normative: This paper is motivated by the normative prior that estimation results should not be determined by the specific imputation rule used to impute

---

[2] Transition periods are affected if they are either followed or preceded by an interregnum period because the authors of Polity decided to first recode interregnum periods and only then to recode transition periods using the previously imputed values of the interregnums.

missing values unless researchers can provide a compelling argument that the specific imputation rule used is optimal. If such an argument cannot be made, researchers should check the robustness of their results to using different imputation rules.

We are hardly the first to argue that imputation by a simple single rule as conducted by the Polity authors can be misleading and – in any case – misrepresents the uncertainty of the estimate. This critique of simple imputation rules along with enhanced computational power of modern computers lead to the rapid increase in the popularity of multiple imputation techniques (Rubin 1976, 1996; Allison 2000; King et al. 2001). In brief, multiple imputation algorithms analyze the estimation model at hand repeatedly, each time using imputed values for missing observations derived from imputation models with random variation. Final coefficients are then averaged over the various iterations with the standard errors coming from the variation of the estimated coefficients across the imputation variations – what Rubin (1976: 585) refers to as sampling distribution inference. This approach has clear advantages over a simple, poorly justified imputation rule. Yet, the flipside of this approach results from the fact that 'models' used for imputation purposes are often poorly specified.

The approach that we employ here accepts the general underlying premise of the multiple imputation approach, namely uncertainty about how to impute missing values. Having this important aspect in common, we part with the traditional multiple imputation approach in two important aspects. First, rather than regarding multiple imputation as the always best approach for dealing with missings, we see it as one additional strategy, complementary to rule- and theory-based single imputation strategies. Thus, we prefer to interpret the different model estimates employing various alternative imputation strategies as robustness checks. If various similarly plausible imputation strategies exist, the parameter estimates of the final model should be robust (not significantly different). If this is not the case, then the choice of the

imputation strategy drives the results. Second, we highlight the importance of using theory to specify the imputation stage of the multiple imputation strategy, simply because the multiple imputation procedure can only be as good as the model from which values are generated. As Gary King (1986: 669) reminded social scientists many years ago: "Some a priori knowledge, or at least some logic, always exists to make selection better than an a-theoretical computer algorithm." We thus use information on the best-performing of our single imputation alternative variables to construct a more theory-based multiple imputation model. We apply our approach to two re-analyses of Fearon and Laitin's (2003) and Krause and Suzuki's (2005) studies of civil war onset to demonstrate the influence of Polity's imputation rule. Re-analyzing these two studies we find that the hypothesis that semi-democracies or anocracies are more likely to experience civil war onset is less robust than previously thought.

2. Polity IV's Coding Rules for Regime Interruption, Transition, and Interregnum years: *polity2*

Polity is based on expert judgment on aspects of institutionalized democracy and autocracy within a country (Jaggers and Gurr, 1995), based on theories of institutions and authority developed by Gurr (1974) and Eckstein and Gurr (1975). Critical evaluations of Polity can be found in, for example, Gleditsch and Ward (1997) and Munck and Verkuilen (2002). In this article, we are exclusively interested in the coding of the *polity2* variable, which is identical to the *polity* variable with the exception of periods of interruption, interregnum and transition. *Interruptions* are periods of foreign occupation as well as periods in which two or more countries became "involved in short-lived attempts at the creation of ethnic, religious, or regional federations" (Marshall and Jaggers 2007: 17). 183 out of 15,036 observations fall into this category (1.2 per cent). *Interregnums* signify the complete collapse of the central

political authority, typically during periods of civil war (179 observations or 1.2 per cent). *Transitions* are periods between two political regimes that substantially differ from each other. They are periods in which "the implementation of generally accepted and substantially altered principles of governance is incomplete and fluid, resulting in mixed patterns that are difficult to define as either those of the old regime or those of the new regime" (Marshall and Jaggers 2007: 18). They account for 316 observations in the data set or roughly 2.1 per cent.

While the *polity* variable flags these periods with values of -66, -77 and -88, respectively, and thus as essentially missing, the *polity2* variable converts these placeholder values based on simple coding rules. We cannot describe the modification leading to *polity2* any better than the codebook, so we quote from there (Marshall and Jaggers 2007: 16):

> It modifies the combined annual POLITY score by applying a simple treatment, or "fix," to convert instances of "standardized authority scores" (i.e., -66, -77, and -88) to conventional polity scores (i.e., within the range, -10 to +10). The values have been converted according to the following rule set:
>
> -66 Cases of foreign 'interruption' are treated as 'system missing'.
>
> -77 Cases of 'interregnum' or anarchy, are converted to a 'neutral' Polity score of '0'.
>
> -88 Cases of 'transition' are prorated across the span of the transition. (…)
>
> Ongoing (-88) transitions in the most recent year (…) are converted to system missing values. Transitions (-88) following a year of independence, interruption (-66), or interregnum (-77) are prorated from the value 0.

Polity2 thus provides *modified* democracy scores during periods of interregnum and transition in addition to the *conventional* democracy scores provided by *polity* for normal times.

3. Consequences of the *polity2* Imputation Rules

Given that there is no disaggregated information in the Polity IV data set on the political regime during periods of interruption, interregnum and transition, the coding rules appear reasonable at first sight: years in which foreign powers exert political control over a country are coded as missing, interregnum years are coded as what the authors of Polity IV deem a 'neutral' score and transition years are linearly interpolated. Yet, on closer inspection, the *polity2* variable turns out to be problematic.

Any good measure has to fulfill two requirements: it has to be reliable and valid (Carmines and Zeller 1979). Reliable here means that in repeated measurements the measured values will be the same or sufficiently similar. With the exception of countries under German occupation during the Second World War, the Polity IV authors seem to have applied their modification rules consistently. We thus do not doubt that repeated measurements would lead to reliable measures. However, we doubt that the coding rules provide a valid measure of democracy during affected country years, i.e. we question whether *polity2* validly measures the level of democracy in these circumstances. This is because on closer inspection the rules for the coding of *polity2* for interregnum and affected transition years give values that are implausible and likely to be misleading regarding the political regime in many of the affected country years.[3] In other words, *polity2* lacks what is called face validity. To persuade readers

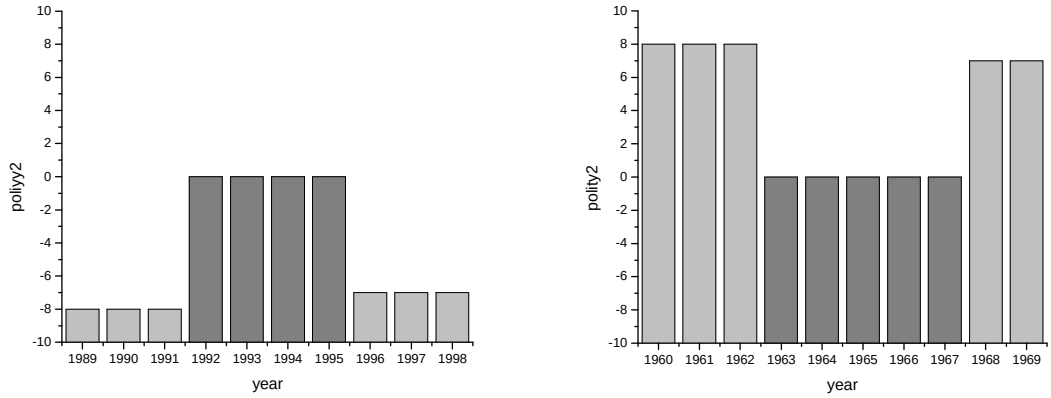
[3] We say *implausible* rather than *wrong* and *likely to be misleading* rather than *misleading* because in the absence of knowledge on the values of the variables underlying the polity score in interregnum and affected transition years it is of course impossible to know what the "true" aggregate democracy score should be.

that this is not just our own idiosyncratic judgment, we will discuss some concrete examples below. On a general level, however, it is not surprising that a mechanic imputation of a score of 0 for all cases of interregnum without using any country-specific information cannot produce a democracy score that commands face validity.

We identify two problems. We dub them the interregnum and the interregnum-cum-transition problem. Note that both problems are caused by the modification of *polity* during interregnum periods. Since the authors of Polity recode interregnum years before they recode transition years, the interregnum recoding rule influences the recoding result of the transition recoding rule if transition years either precede or follow interregnum years. We discuss both problems separately.

The Interregnum Problem

The interregnum problem emerges from the coding of interregnum periods with a value of zero, which is meant to represent a *neutral* regime score. However, just because 0 lies in the middle between -10 and 10 does not make this a democracy score that is any more neutral than 0 degrees Celsius or Fahrenheit would represent a neutral temperature. Depending on the political regime before and after the interregnum period, a value of zero can represent everything but a neutral regime. To illustrate our argument, figures 1a and 1b compare the *polity2* scores of Afghanistan over the period 1989-1998 to those of Cyprus over the period 1960-1969.



Note: interregnum years are indicated by dark grey bars

Figures 1a and 1b: *polity2* scores of Afghanistan (left) and Cyprus (right)

The two figures reveal that the coding rule for interregnum years can generate two rather different effects. First, for some countries with an undemocratic regime before and after interregnum (here exemplified by Afghanistan), the interregnum years are the most democratic years these countries have experienced over a certain time period. In other words: for Afghanistan and similar cases, an interregnum marks an improvement in the level of democracy. In the case of Afghanistan, if we trusted *polity2*, then the country would have gone through a much more democratic period during the four years of civil war than under the Communist rule of Najibullah before or the Taliban-regime after the interregnum. If we compare the *polity2* scores to the scholarly literature on Afghanistan's political regime, then for most if not all observers, *polity2* does not accurately portray the state of Afghan democracy during the years 1992 to 1995 (Rubin 2000; Marsden 1998). Such likely mis-representation is not restricted to one isolated case. Instead, for example, along similar lines, Czechoslovakia in 1968, Ethiopia in 1974, Cambodia in 1975, Afghanistan in 1978, Nicaragua from 1979 to 1980 as well as Somalia from 1991 onwards are coded as much more democratic during years of social unrest and civil war than either before or after. While the

coding of Czechoslovakia is perhaps broadly in line with historical accounts of the Prague Spring (Williams 1997), the other cases are clearly at odds with common perceptions. In fact, for Afghanistan and Somalia, state collapse and anarchy represent by far the most democratic periods in their entire post-independence history. The fundamental problem is that when a country has an undemocratic regime before and after an interregnum period, then a democracy score of zero is unlikely to represent a *neutral* democracy score.

The *polity2* rules create the opposite effect in the case of Cyprus. Polity IV interprets the years 1963 to 1967 as interregnum years, even though neither the Correlates of War (Sarkees 2000) nor the Uppsala/PRIO (Gleditsch et al. 2002) data sets indicate an armed civil conflict. Whatever the reasons behind the decision to treat this period as interregnum, the effect is to make Cyprus far less democratic than either before or after these years. In fact, the allegedly neutral value of zero during the period 1963 to 1967 amounts to the country's most undemocratic period in its entire post-independence history. Similarly, in Sierra Leone during 1997 to 2001, the Solomon Islands during 2000 to 2002 and Comoros in 1995 civil unrest prompted the Polity IV authors to code an interregnum, leading to a decline in the *polity2* score relative to previous years. For Sierra Leone with its brutal civil war the *polity2* score may not be far off the true state of democracy during this period (Zack-Williams 1999). The same may apply to the Solomon Islands, which slipped into chaos after rebels kidnapped Prime Minister Bartholomew Ulufa'alu in June 2000 (Reilly 2004). But for Comoros in 1995 the *polity2* score is likely to be misleading since its semi-democratic regime did not become any more autocratic when French troops helped to thwart a military coup led by Bob Denard (Anonymous 1998).

Similar to the other cases discussed previously, the (reverse) problem is that when a country has a democratic regime before and after an interregnum period, then a score of zero

is unlikely to represent an appropriate level of democracy either. Comparing both types of cases, we find that according to *polity2* an interregnum increases the level of democracy in autocracies and reduces the level of democracy in democracies, an outcome we do not find plausible.

Making things worse, the problematic *polity2* coding rules are not applied consistently, as the countries occupied by Germany during the Second World War show. To recall, foreign occupation should lead to a *polity* score of -66 and thus to a system missing value for the *polity2* variable. However, for no clear reason, the Polity IV project codes Austria, Czechoslovakia, Poland, Albania, Greece and Bulgaria as interregnums during their period of German occupation (thus their *polity2* score is 0), whereas Belgium, Denmark, the Netherlands and Norway are coded as interruptions (thus their *polity2* score is missing).[4]

Without doubt, periods of interregnum are difficult to code. However, we contend that the authors of Polity IV made inconsistent and implausible decisions when facing this difficulty. Coding all interregnum years as zero on the assumption that this represents a politically *neutral* value is naïve for basically three reasons. First, the political processes which cause the coders of Polity IV to treat country-years as interregnum periods vary largely. Both moderate political turbulence and unrest as in Cyprus, civil wars as in Afghanistan, and the German

---

[4] The case of Austria is particularly problematic. In 1938, German troops invaded Austria to bring its South-Eastern neighbour 'Heim ins Reich' ('home into the empire'), which was the Nazi-euphemism for the annexation of the country. After annexation, Austrians had the same rights (or lack thereof) and the same obligations and were subjected to the same political regime as Germans. However, while Germany has a *polity2* score of -9 from 1938 onwards, Austria is coded as 0. This is inconsistent because either Austria was occupied by a foreign power or experienced a short-lived attempt at building an ethnic federation, in which case its *polity2* score should be set to missing, or Austrians shared the same fate as Germans, in which case its *polity2* score should be set to -9.

occupations of Poland, Czechoslovakia, Albania, Yugoslavia, Greece and Bulgaria are equally treated as interregnum years and assigned a score of zero, which we contend in most cases provides implausible information on the state of political regime in these countries during these years. Second, treating interregnum periods equally without regard to the political regime context leads to arbitrary dynamics in the level of democracy. Interregnum periods bring about an increase in the level of democracy in most countries, while in others the level of democracy declines during periods of anarchy. Third, if *polity2* is used in regression analysis, then coding interregnum years as zero is not "neutral" either because the estimation of the coefficients of the *polity2* as well as in principle of all other explanatory variables correlated with *polity2* are affected. Thus, the coding of 0 for periods of interregnum is potentially not "neutral" to causal inferences based on regression analysis either – a point to which we come back later.

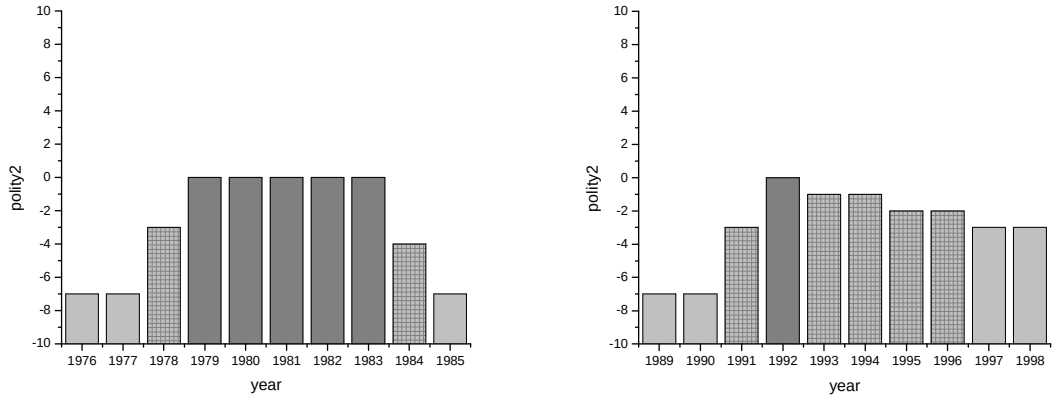
The Interregnum-cum-Transition Problem

The lack of validity in the *polity2* variable caused by the coding of interregnum periods potentially worsens if such periods are preceded or followed by periods of transition.[5] The interregnum-cum-transition problem emerges because the Polity IV authors decided to first code the *polity2* score for the interregnum years and only then to code the transition years. As a consequence, any problematic coding of interregnum years carries over to the coding of transition years if a transition follows an interregnum or vice versa. In Polity IV, 128 observations are affected by the interregnum-cum-transition problem.

---

[5] Note that we see no reason to alter the re-coding rule employed by the Polity project for pure transition periods – that is, transition periods which are neither preceded nor followed by an interregnum period.

To illustrate our arguments, figures 2a and 2b plot the *polity2* scores for Chad and Angola over the periods 1976 to 1985 and 1989 to 1998, respectively.



Note: interregnum years indicated by dark grey bars, and transition periods by patterned bars

Figure 2a and 2b: *polity2* Score of Chad (left) and Angola (right)

As in the case of Afghanistan, we find that the coding rules underlying the *polity2* variable lead to a sudden increase in the democracy score of both countries during interregnum years (see Azevedo 1998; Brittain 1998). The problem resulting from assigning a value of zero to Chad during 1979 to 1983 and to Angola in 1992 is potentially worsened by interpolating from this value, which results in an artificial upward movement in the *polity2* score for Chad in 1978 and for Angola in 1991 and an artificial downward movement in the *polity2* score for Chad in 1984 and Angola in 1993 to 1996. Similar cases in the period after the Second World War include Burundi from 1991 to 1996, Congo from 1962 to 1965, Cuba from 1958 to 1961, Guinea-Bissau from 1997 to 2000, Laos from 1960 to 1975, Lesotho from 1997 to 2002, Liberia from 1989 to 1997, Pakistan from 1968 to 1973 and Sierra Leone from 1996 to 2002.

4. Existing Strategies Available to Researchers with *polity2*

With the emergence of *polity2*, researchers using Polity can choose any of at least four existing strategies for their research designs. First, they can stick to the old *polity* variable and thereby listwise exclude -66, -77, and -88 observations – thus accepting that the democracy score for these types of regimes is impossible to identify. Second, they can use *polity2* and thereby accept the imputations to *polity* undertaken by the authors of Polity IV on the basis of their coding rules. Third, they can employ the *polity2* variable plus adding, in multivariate regression analysis, a set of three dummy variables: one for values of -77 in countries that have negative *polity2* values before and after interregnum, one for values of -77 in countries that have positive *polity2* scores before and after interregnum, and one for -88 observations. Fourth, they can apply a Heckman selection model which treats interruption, interregnum, and transition years as non-selected in the first stage and then accounts for the non-selection hazard in the second stage of their estimations.

Each of these options is problematic. Listwise exclusion generates a non-random selection criterion which will often be correlated with many variables included in the estimations. As a consequence, in many cases listwise exclusion will lead to biased estimates (Honaker and King 2010). Using *polity2* is problematic because its scores lack face validity, as argued in the previous section, and, more importantly, may impact causal inference as argued in section 7. Using *polity2* with correction via dummy variables for interregnums and affected transition periods should give less biased, but also less efficient estimates if the dummy variables are correlated with variables of interest. Also, an identification problem occurs because it is impossible to interpret the effect of these dummy variables as either the effect of interregnums and affected transition periods per se or as correction for the coding rule applied by the authors of Polity IV. The Heckman selection model is not very

problematic from an econometric point of view. But it requires the development of a theoretical model for the first, the selection stage, i.e. a model which can explain interruption, interregnum and transition periods. Such a task appears to be difficult, if not impossible. Given this set of rather unsatisfactory existing options, we will now suggest alternatives.

5. Alternative Imputation Strategies

In this section we discuss three methods that provide alternative imputation strategies to Polity's coding of *polity2*. One method uses no information from outside the Polity IV data set, but applies different coding rules to impute values for periods of interregnum and affected transitions. The second method employs outside information. Specifically, we generate out-of-sample predictions for interregnum years by using another measure of democracy (from the Freedom House data set) and theoretical determinants of democracy. For both methods, we then use the out-of-sample predictions for interregnum years to re-code those transition years that either follow or precede interregnum years. The increasingly popular method of multiple imputation, already mentioned in the introduction, provides a third solution to the imputation problem.

Recoding According to Different Rules

We suggest three alternative coding rules for interregnum and affected transition years that do not depend on any information from outside the Polity IV data set. All three start from *polity2*, but set all observations during interregnum years and all transition years that immediately follow or precede interregnum periods to missing.

Rule 1 (minimum level). Set the interregnum years to the lower of the two polity scores bordering the interregnum period. Then use linear interpolation to add the affected transition years (we dub this variable *polity2min*).

Rule 2 (interpolation). Use linear interpolation to fill in both interregnum and interregnum-cum-transition years (*polity2inter*).

Rule 3 (maximum level). Set the interregnum years to the higher of the two polity scores bordering the interregnum period. Then use linear interpolation to add the affected transition years (*polity2max*).

In all three rules, interpolated values are rounded to the nearest integer in case the interpolated value is not an integer. Of course, the choice of these three recoding rules is somewhat arbitrary. However, they have three advantages. First, contrary to the mechanic *polity2* coding rules, they use country-specific information immediately preceding and/or following the period of interregnum. Second, they therefore do not treat interregnums in countries with otherwise low or high levels of democracy equally. Third, they constrain the level of democracy to lie within the range before and after the interregnum. Even though exceptions are of course possible, in most cases it seems plausible that interregnums were neither less nor more democratic than the regime either before or after this period.

Out-of-sample Predictions

Alternative measures of democracy are likely to be powerful candidates for out-of-sample predictions. Of the data sets reviewed and evaluated in Munck and Verkuilen (2002), the majority is unsuitable due to poor country coverage or because they only provide a dichotomous regime type measure. Vanhanen's (2000) data set has very large spatio-temporal coverage, but we doubt whether regime type can be measured solely based on election results

data, which not only ignores important institutional features of democracies, but also precludes a change in the democracy score between elections.

The one suitable candidate for our purposes is the Freedom House data set. Covering the period since 1972, it is nowhere near as comprehensive in the temporal domain, but it covers all countries that are also contained in the Polity IV data set.[6] The use of Freedom House data for research spanning a period of time is not unproblematic since its scale changes slightly over time and it was not originally designed as a time-series. However, the authors of Freedom House and of Polity use similar definitions of democracy and both are based on expert judgments. And even though the indicated level of democracy in certain countries can differ between the two sources – for example, Freedom House considers Qatar and Kuwait as more democratic than Polity IV, whereas the reverse is true for Russia – the correlation of the two democracy scores is very high. Given that Freedom House is the second most used data set for measuring democracy after Polity and since it provides values for all country years since 1972 that are coded as interregnum and transition in Polity, using Freedom House data for out-of-sample prediction of the missing polity scores is an obvious choice.

In lieu of or in addition to the Freedom House values, one can use theoretically informed variables commonly regarded as determinants of democracy. Our choice of explanatory variables is informed by the number of missing observations. Specifically, we exclude variables with poor spatio-temporal availability. Analyses of the determinants of democracy typically find the level of economic development (Lipset 1959; Burkhart and Lewis-Beck

---

6 www.freedomhouse.org. In fact, it covers all countries in the world and therefore a good deal more than Polity, which restricts itself to countries with population size above half a million. In principle, one could apply our method to create out-of-sample predicted values to generate imputed Polity scores for states with small population sizes, but we do not do so here since our only concern lies with the *polity2* recoding rules.

1994; Przeworski and Limongi 1997; Barro 1999) and neighborhood effects (Gleditsch and Ward 2006) to exert the largest effect on the level of democracy. Accordingly, our theoretical model includes per capita income and its squared term, with data taken from Gleditsch (2002), as well as the lagged mean value of democracy in neighboring countries within a radius of 500 kilometers. To account for some heterogeneity across countries, we also include a dummy variable which is coded 1 if a developing country shares the main language with a developed country, a dummy variable for former Western colonies, dummy variables for regions as per World Bank classification as well as a dummy for countries with a predominantly Muslim population.

To generate new variables based on out-of-sample predictions, we start with the *polity* variable and, first, set interruption, interregnum and interregnum-cum-transition periods to missing. Using an ordered logit estimator, we regress this variable on three sets of explanatory variables: once only the determinants of democracy (*polity2* (det)), once only Freedom House's political rights and civil liberties measures (*polity2* (FH)) and, finally, once combining these two sets of variables (*polity2* (FH & det)).[7] We then reset all interruptions back to missing, replace all interregnum years by the rounded predicted values of our estimations and fill the interregnum-cum-transition years by linear interpolation (rounding to the nearest integer if necessary).[8]

---

[7] Adding 10 to the revised polity variable to create a strictly non-negative categorical variable and then employing the negative binomial regression model leads to very similar out-of-sample predictions for interregnum years (correlated at .97 with the ones from the ordered logit model). We found that the ordered logit model has a small advantage over the negative binomial model in terms of goodness of fit of the predicted values with the in-sample *polity* observations.

[8] To construct predicted values we multiply the predicted probabilities of each polity category with the respective polity value and then round it to the nearest integer.

Multiple imputation

Both our recoding rules and our out-of-sample predictions generate single imputed values. An alternative is provided by the technique of multiple imputation. In a first, multiple imputation stage, a specified number M of imputed values are created based on some imputation model. Then, in the estimation stage, the estimation is performed on each of the M imputations and final results are typically obtained pooling the M set of multiple imputation estimates. We will discuss multiple imputation in more detail in section 8, where we use it in our re-analysis exercise and where we propose to employ information on the best-performing single imputation variables in the imputation stage this strategy. Before we do so, we use a relatively simple out-of-sample prediction technique to evaluate the performance of our theory-based single imputed value variables based on recoding and out-of-sample predictions.

6. An Evaluation of the Alternative Single Imputation Polity Scores

The imputation methods described above generate six new single imputation variables, three for the rule-based method that uses no information outside Polity IV and three for the out-of-sample predictions using outside information. One research design option would simply be to use all these alternatives to *polity2* in robustness tests. However, the different single imputation variables are of varying quality, so it appears to be good praxis to narrow down the set of imputation rules used in a robustness test. For this reason, this section analyzes the quality of the different single imputation variables based on out-of-sample techniques. In section 7, where we provide a re-analysis of two studies of civil war onset, we then use the best-performing single imputation variables, use a standard multiple imputation model as well as a combination of the two techniques, in which we employ the information on the best-

performing single imputation variables to construct a more theory-based multiple imputation model.

To start, table 1 provides a simple correlation matrix of the seven new variables and *polity2*. Clearly, the correlation between *polity2* and our recoded variables are close to one due to the small number of missings, the correlation between *polity2* and *polity2* (det) is rather low, while the correlations with *polity2* (FH) and polity2 (FH & det) are very high.

Table 1. Correlation matrix for *polity2* and alternative measures (N = 4519).

	polity2	min	inter	max	(det)	(FH)
polity2min	0.995					
polity2inter	0.997	0.999				
polity2max	0.998	0.998	0.999			
polity2 (det)	0.711	0.711	0.711	0.710		
polity2 (FH)	0.916	0.918	0.919	0.918	0.702	
polity2 (FH & det)	0.930	0.931	0.932	0.931	0.772	0.983

Since the true level of democracy is unknown for interregnum years, we cannot establish *a priori* which of the six alternatives is superior. In the following we will assume that procedures which predict values that are closer to the conventional *polity2* scores during normal times also produce superior values for years of interregnums. We admit that this is a strong assumption. Using it, however, we can establish a ranking of the three variables based on out-of-sample predictions, while we cannot apply this criterion to the variables derived from our re-coding rules since by design they would be perfectly collinear to *polity2*.

Table 2. Goodness-of-Fit Tests between Out-of-Sample Predictions and Conventional Polity Scores (N varies)

	correlation	regression (R^2)	regression (RMSE)
	Dependent variable: conventional polity scores		
	higher is better	higher is better	lower is better
polity2 (det)	0.67	0.45	5.70
polity2 (FH)	0.92	0.85	2.92
polity2 (FH & det)	0.93	0.87	2.74

The emerging picture from table 2 is clear: using another measure for democracy like the Freedom House data predicts conventional *polity2* scores much better than using a theoretically informed democracy model according to all three criteria (correlation, pseudo R^2 and root mean square error). Adding the theoretical determinants as additional regressors to the Freedom House variables improves the goodness-of-fit further, but only marginally. Hence, if our assumption that an imputation rule is better the better it predicts conventional *polity* scores is correct, then the variables based on out-of-sample predictions using the Freedom House data together with the theoretical determinants of democracy performs best. We call this variable *polity2pred*, discarding the other two variables as they are inferior to *polity2pred*.

We can now use the *polity2pred* variable to evaluate the rule-based imputations and the mean of the multiply imputed values. Similar to before, we assume that the rule whose imputed values fit more closely with *polity2pred* during periods of interregnum and affected transitions is superior. Table 3 displays the results from the same set of goodness-of-fit tests as we already used in table 1, to compare *polity2min*, *polity2inter*, and *polity2max* with the out-of-sample-predictions of *polity2pred* for interregnum and affected transition periods.

Table 3. Goodness-of-Fit Tests between Out-of-Sample Predictions and Rule-Based Imputations for Interregnum and Affected Transition Years (N = 39)

	correlation	regression (R^2)	regression (RMSE)
	Dependent variable: Out-of-sample predictions from model using Freedom House data and determinants of democracy		
	higher is better	higher is better	lower is better
polity2min	0.83	0.67	2.23
polity2inter	0.80	0.64	2.35
polity2max	0.77	0.58	2.54

Apparently, the minimum level rule performs better than the interpolation rule, which in turn performs better than the maximum level rule.

7. Improving Inferences by Combining Theory-based Single Imputation with Multiple Imputation: Two Re-analysis Examples

We have argued above that the Polity IV project's coding rules for interregnum and affected transition years lead to democracy scores that lack face validity and have proposed several alternative variables. Given that country years of interregnum and interregnum-cum-transition represent a very small share of the total observations and replacing *polity2* with one of our alternative variables may thus only affect a small proportion of a sample used in research, the question is whether scholars must be concerned about the manifest problems with the *polity2* variable. Does the lack of validity in *polity2* affect either descriptive or causal inference?

If researchers are mainly interested in global population descriptives (e.g., changes in the global mean of democracy), then the affected share of observations is so small that our proposed alternative measures would be very unlikely to make a significant difference. If, however, they are interested in sub-sample population description (e.g., the number or share of autocracies, democracies or anocracies, that is countries in between autocracy and democracy), then descriptive inference can be affected as *polity2* would often code affected

country years as anocracies whereas our alternative variables would often code them as either autocracies or democracies.

Most researchers use the Polity data for causal inference, however. We therefore explore the effect of using our alternative variable on causal inference in greater detail. Researchers should indeed be concerned about the problems that *polity2* can create for several reasons. First, using more accurate information is a value in itself. Social science data are often noisy, which renders estimations inefficient. More valid measures lower measurement error and thus lead to more reliable estimation results.

Second, using the *polity2* variable can lead to significantly different estimates compared to using each of our alternative variables. This is more likely to happen the larger the share of interregnum and interregnum-cum-transition country years in the sample. Examples consist of a pre-First World War sample of predominantly Latin American countries, a sample of predominantly European countries during or around the Second World War, or a post-Second World War sample of mainly Asian and Sub-Saharan African countries.

Third, even global samples can be affected if scholars transform the *polity2* variable into three dummy variables for autocracies, democracies and semi-democracies (sometimes referred to as anocracies) or test for a non-linear effect in the continuous *polity2* variable. Interregnums will always be categorized as periods of anocracies in *polity2* (as will some affected transition years), whereas our alternative variables will sometimes categorize them as autocracies, anocracies or even democracies, depending on the case, thus changing the distribution and means of covariates in these three groups of countries.

Fourth, other global samples can also be affected if scholars employ a fixed effects or a first differences model. In these cases, the between-variation of the variables is ignored so that the within-variation in the *polity2* variable becomes the sole information used in the

estimation. Accordingly, the large upward and downward changes in *polity2* that often happen during interregnum and interregnum-cum-transition years become an issue and are likely to exert a large influence on the estimates.

Fifth, any of the causes of wrong inference discussed so far will become exacerbated if the dependent variable is either state failure (collapse of central political authority) itself or a determinant of state failure. These events make it very likely that a country will be coded by Polity IV as going through an interregnum. In other words, the dependent variable determines the coding of *polity2*, hence using *polity2* as an explanatory variable will lead to biased estimates due to endogeneity. Thus, studies of civil war and state collapse are more likely to find that country years with a *polity2* score of zero have a higher likelihood of civil war and state collapse because civil wars and state collapse to some extent cause a *polity2* score of zero, rather than the other way around.

To illustrate some of the potential problems with inference based on *polity2*, we re-analyze two empirical analyses of civil war onset using a) our best-performing single imputation variables; b) a standard multiple imputation model; and c) a combination of the two in which information on the best-performing single imputation variables is used to construct a more theory-based multiple imputation model. One of these studies, Krause and Suzuki (2005), was selected because it employs two separate samples for Asian and Sub-Saharan African countries over the period 1950 to 1992, which according to our discussion above should make problems with statistical inference using *polity2* more likely. The other, Fearon and Laitin's (2003) prominent study of civil war onset, was chosen to show that such problems can occur, if to a far lesser extent, even in a global sample over the period 1945 to 1999. Both Fearon and Laitin (2003) and Krause and Suzuki (2005) find that anocracies are more likely to experience civil war onsets than either autocracies or democracies.

Fearon and Laitin define anocracies as countries with a *polity2* score between -5 and 5 while democracies are countries with a *polity2* score between 6 and 10. Political instability is a dummy variable set to one if there has been a 3-point or more change in *polity2* in the previous three years or a period of interregnum or transition. Model 4.1 of table 4 replicates model 3 of table 1 from Fearon and Laitin (2003), which is their preferred model. The anocracy dummy variable is positive and statistically significant at $p = 0.028$, whereas the democracy dummy variable is not. This implies that anocracies have a statistically significantly higher risk of civil war onset compared to autocracies, the omitted reference category, whereas democracies do not. Note, however, that the anocracy and democracy dummy variables are not statistically significantly different from each other.

In models 4.2 and 4.3 we replace the anocracy and democracy dummies, which were based on *polity2*, with similarly constructed dummy variables for anocracies and democracies derived from *polity2min* and *polity2pred*, respectively, the two best-performing single imputation variables from section 6. The anocracy dummy, which was significant at the 5 per cent level in Fearon and Laitin (2003) is no longer statistically significant if *polity2min* or *polity2pred* are used. These changes in the estimated coefficients and the standard errors result from the re-classification of several observations from anocracies in *polity2* to autocracies in *polity2min*.

Model 4.4 uses a multivariate normal regression multiple imputation technique based on Bayesian iterative Markov chain Monte Carlo (MCMC) procedures to impute the 66 interregnum and affected transition observations, using all the variables in the estimation

model plus regional dummy variables to create 100 sets of imputed values.[9] We call this a standard multiple imputation model. The coefficient of the anocracy dummy variable is again no longer statistically significant at the 10 per cent level.

In model 4.5, we repeat the MCMC multiple imputation procedure, but this time specifying the imputation stage differently. Honaker and King (2010) argue that standard imputation models often work very poorly for time-series cross-section data as such data violate the assumptions of conditional independence and exchangeability of observations. They suggest, among other things, to include lags and leads of the variables that have missings as variables in the imputation model in order to impose smoother trends in the variables to be imputed than what is typically generated by a standard imputation model. One of the two best-performing variables from section 6, *polity2min*, allows us to do exactly that (we additionally include *polity2max* to follow their recommendation to include leads in addition to lags). Honaker and King (2010) also propose to incorporate expert knowledge in the form of new types of Bayesian priors. We agree with the idea of including expert knowledge, but do so in the form of including variables based on expert knowledge in the imputation stage. In section 6, the out-of-sample predictions based on Freedom House data proved to be a good predictor of *polity* and since the Freedom House data can be regarded as reflecting expert knowledge, we also include them in the imputation stage (after 1972).[10] With

---

9 We convert the continuous imputed values to discrete values of 0 or 1 to be consistent with the dummy variable approach of Fearon and Laitin (2003). Using a logit imputation technique instead leads to similar results.

10 In principle, one could include further variables such as the theoretical determinants of democracy in the imputation stage, but since these variables themselves have missings and, as section 6 showed, did not add much to the out-of-sample predictions based on Freedom House data alone, we do not include them here.

this alternative multiple imputation model, the anocracy dummy variable is now statistically significant at the 5 per cent level. If one accepts that, being theory-based, this multiple imputation model 4.5 is superior to model 4.4, then the analysis would again corroborate the hypothesis that anocracies are more prone to experience civil war onset.

Model 4.6 repeats the MCMC imputation, but this time combines the variables from the respective imputation stages of models 4.4 and 4.5. In other words, the imputation stage now includes all the variables of the estimation model plus the *polity2min*, *polity2max* and the Freedom House variables. The estimation results are very similar to the ones from model 4.5, which suggests that the variables from the estimation model add very little information to the imputation stage. We take this as evidence suggesting that imputing the interregnum and affected transition years exclusively with *polity2min*, *polity2max* and Freedom House variables is sufficiently good for making valid inferences. In fact, if one were to run a standard logit estimation model in which one included the average of the 100 imputed polity values of the imputation stage from model 4.5, then results are very, very similar to the ones from the multiple imputation model 4.5 (results not shown). We take this as evidence that if researchers do not want to undertake multiple imputation based on our *polity2min*, *polity2max* and Freedom House variables themselves, then they can safely use the mean of imputed *polity* values from an imputation model, in which these variables were used.[11]

Table 5 exactly repeats this exercise for the estimation results from table 1 of Krause and Suzuki (2005) for their sample of Asian countries. Whereas Fearon and Laitin use anocracy

---

[11] Employing the mean of imputed values instead of using multiple imputation will slightly under-estimate standard errors as one single value is used instead of the multiply imputed values, but due to the relatively small share of imputed values, there is almost no difference to the standard errors that would result from employing multiple imputation.

and democracy dummies, Krause and Suzuki employ the continuous *polity2* score (to which they add 10 to make it strictly non-negative) and its squared term. They find a non-linear effect in both samples: *polity2* is positive and statistically significant, whereas its squared term is negative and statistically significant, which again suggests that the risk of civil war onset is highest in countries with an intermediate *polity2* score, i.e. in semi-democracies or anocracies (model 5.1). Estimating, with one small exception[12], the same set of models (in the same order) as in the replication of Fearon and Laitin (2003), we find the evidence for the anocracy hypothesis to be non-robust. The coefficients of the regime and regime squared variables become less significant or even insignificant when our single imputed alternative variables are used and become altogether insignificant in any of the multiple imputation exercises. We have also replicated Krause and Suzuki's estimations for their Sub-Saharan Africa sample (re-analysis results not shown). Neither regime nor regime squared is significant in any of the models that does not use *polity2*.

The results presented in Krause and Suzuki (2005) and, to a far lesser extent, in Fearon and Laitin (2003) are thus sensitive to our recoding of the *polity2* variable or to using multiple imputation for interregnums and affected transition periods.[13] In *polity2* all interregnums and some of the affected interregnum-cum-transition periods are coded as anocracies. Moreover, civil war onset, the dependent variable, is correlated with the likelihood that a country experiences a period of interregnum and interregnum-cum-transition since civil wars are the prime cause of interregnums. Thereby, the *polity2* score inflates the probability that a study of

---

12 The maximum likelihood did not converge when using *polity2pred*. We therefore used the out-of-sample predictions based on the theoretical determinants of democracy instead.

13 For additional evidence that the anocracy result in Fearon and Laitin (2003) is fairly robust to changes in research design, see De Soysa and Neumayer (2007).

civil war onset will find that anocracies have a higher likelihood of civil war onset. In sum, the evidence for the anocracy-hypothesis seems to be less robust than previously believed.

8. Conclusion

Political system attributes in general and the level of democracy in particular are difficult to measure. This holds true even more so in difficult times – years in which the execution of political power is hampered by civil wars, coups, social unrest, foreign intervention, leadership struggles and similar events. It was thus a good idea of the makers of the Polity project to flag these years, which allows researchers to give them special attention. It was not an equally good idea, however, to apply a seemingly plausible, but in fact problematic fix for providing a *polity2* score for years of interregnum and transition. A polity score of zero is not a "neutral", equally suitable score for all instances of interregnum. Neutrality in an arbitrarily chosen scale does not exist. For most countries experiencing interregnums, these years were coded as more democratic than either before or after interregnum. Similarly, for some other countries experiencing interregnums, these years were coded as less democratic than either before or after interregnum. We have argued that the *polity2* coding rules produce values that lack face validity.

We have briefly evaluated existing strategies available to researchers, but found all of them problematic. We therefore used three methods to construct alternative variables to *polity2*. We discussed several general conditions under which using our alternative variables would change causal inference compared to *polity2*. For Krause and Suzuki (2005), we have shown in a re-analysis that applying these alternative variables or using multiple imputation techniques leads to different statistical inferences concerning the effect of anocracy on the risk of civil war onset. The results of Fearon and Laitin (2003), however, appear to be more robust

if our theory-based single imputation values are in turn included in the multiple imputation model.

What then should applied researchers do? We recommend that scholars test whether their results derived from using *polity2* are robust to using our alternative variables or using multiple imputation techniques instead. If results turn out to be robust, then causal inference does not seem to suffer much from the problems of *polity2*. If results differ significantly then researchers need to theoretically justify the use of either *polity2* or one of our alternative variables or of multiple imputation. When researchers choose to employ multiple imputation, we also recommend including, additionally or exclusively, our alternative variables plus the Freedom House variables in the imputation model.

Our analysis has wider implications as well, however. What we propose here is a method that combines strategies from theory-based single imputation with the use of multiple imputation algorithms together with a call for robustness checks across estimates from the range of values generated by different imputation methods. We believe that the combination of this method plus robustness checks offers an important step forward not only for dealing with interregnum periods in the Polity data set, but for dealing with missings in general.

Appendix A: Examples of Interregnum and Interregnum-cum-Transition Periods

Table A1 shows the scores of *polity*, *polity2*, *polity2pred*, *polity2min*, *polity2inter* and *polity2max* for Afghanistan over the period 1989 to 1997. Afghanistan is remarkable because the Polity IV coding rules for *polity2* provide the country with a much higher level of democracy than either before or after. In contrast, the four alternative variables suggest a far lower level of Afghan democracy during this time period.

Table A1: Imputed Polity scores for Afghanistan 1989-1998

year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
polity	-8	-8	-8	-77	-77	-77	-77	-7	-7	-7
polity2	-8	-8	-8	0	0	0	0	-7	-7	-7
polity2pred	-8	-8	-8	-6	-8	-8	-8	-7	-7	-7
polity2min	-8	-8	-8	-8	-8	-8	-8	-7	-7	-7
polity2inter	-8	-8	-8	-8	-8	-7	-7	-7	-7	-7
polity2max	-8	-8	-8	-7	-7	-7	-7	-7	-7	-7

Conversely, for Cyprus our alternative variables suggest a far higher level of democracy during the interregnum period (table A2).

Table A2: Imputed Polity scores for Cyprus 1960-1969

year	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
polity	8	8	8	-77	-77	-77	-77	-77	7	7
polity2	8	8	8	0	0	0	0	0	7	7
polity2pred	8	8	8	..	..	..	..	..	7	7
polity2min	8	8	8	7	7	7	7	7	7	7
polity2inter	8	8	8	8	8	8	7	7	7	7
polity2max	8	8	8	8	8	8	8	8	7	7

Table A3 presents the values for Chad over the period 1976 to 1985. Imputations based on out-of-sample predictions and our rule-based recodings lead to similar results: The openness

of Chad's political system has not changed much during the interregnum and affected transition period.

Table A3: Imputed Polity scores for Chad 1976-1985

year	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
polity	-7	-7	-88	-77	-77	-77	-77	-77	-88	-7
polity2	-7	-7	-3	0	0	0	0	0	-4	-7
polity2pred	-7	-7	-6	-6	-8	-8	-8	-8	-7	-7
polity2min	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7
polity2inter	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7
polity2max	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7

Table A4 provides an example how the various alternative variables can lead to somewhat different democracy scores, here for Angola's interregnum-cum-transition period in the 1990s:

Table A4: Imputed Polity scores for Angola 1989-1998

year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
polity	-7	-7	-88	-77	-88	-88	-88	-88	-3	-3
polity2	-7	-7	-3	0	-1	-1	-2	-2	-3	-3
polity2pred	-7	-7	-6	-6	-5	-5	-4	-4	-3	-3
polity2min	-7	-7	-7	-7	-7	-7	-7	-7	-3	-3
polity2inter	-7	-7	-6	-6	-5	-4	-4	-3	-3	-3
polity2max	-7	-7	-3	-3	-3	-3	-3	-3	-3	-3

References

Allison, Paul. D. 2000. Multiple Imputation for Missing Data. A Cautionary Tale. *Sociological Methods and Research* 28, 301-309.

Anonymous, 1998. Comoros. *Africa Research Bulletin* 35:11, 13317-13318.

Azevedo, Mario Joaquim. 1998. *Roots of Violence. A History of war in Chad*. Routledge, Oxford.

Barro, Robert J., 1999. Determinants of Democracy. *Journal of Political Economy* 107:6, S158-S183

Brittain, Victoria, 1998. *The Death of Dignity. Angola's Civil War*. Pluto Press, London.

Burkhart, Ross E. and Lewis-Beck, Michael S., 1994. Comparative Democracy. The Economic Development Thesis. *American Political Science Review* 88:4, 903-910.

Carmines, Edward G. and Richard A. Zeller, 1979. *Reliability and Validity Assessment*. London: Sage.

Casper, Gretchen and Claudiu Tufis, 2003. Correlation versus Interchangeability. The Limited Robustness of Empirical Findings of Democracy using Highly Correlated Data. Political Analysis 11: 2, 196-203.

De Soysa, Indra and Eric Neumayer, 2007. Natural Resources and Civil War: Another Look with New Data. *Conflict Management and Peace Science* 24:3, 201-218.

Eckstein, Harry, and Ted Robert Gurr, 1975. *Patterns of Authority: A Structural Basis for Political Inquiry*. New York: Wiley-Interscience.

Fearon, James D. and Laitin, David D. 2003. Ethnicity, Insurgency and Civil War. *American Political Science Review* 97:1, 75-90.

Gleditsch, Kristian S. 2002. Expanded Trade and GDP Data. *Journal of Conflict Resolution* 46: 712-24.

Gleditsch, Kristian S. and Ward, Michael, 1997. Double Take. A Reexamination of Democracy and Autocracy in Modern Politics. *Journal of Conflict Resolution* 41:3, 361-383.

Gleditsch, Kristian S. and Ward, Michael, 2006. Diffusion and the International Context of Democratization. *International Organization* 60: 4, 911-933.

Gleditsch, Nils Petter, Wallensteen, Peter, Erikson, Mikael, Sollenberg, Margareta and Strand, Håvard, 2002. Armed Conflict 1946-2001. A New Dataset. *Journal of Peace Research* 39: 5, 615-637.

Gurr, Ted Robert, 1974. Persistence and Change in Political Systems, 1800-1971. *American Political Science Review* 68:4, 1482-1504.

Honaker, James and Gary King. 2010. What to Do about Missing Values in Time Series Cross-Section Data. *American Journal of Political Science*, forthcoming.

Jaggers, Keith and Gurr, Ted Robert, 1995. Tracking Democracy's Third Wave with the Polity III Data, *Journal of Peace Research* 32:4, 469-482.

King, Gary, Honaker, James, Joseph, Anne and Scheve, Kenneth. 2001. Analyzing Incomplete Political Science Data. An Alternative Algorithm for Multiple Imputation. *American Political Science Review* 95:1, 49-69.

King, Gary, 1986. How no to lie with Statistics. Avoiding Common Mistakes in Quantitative Political Science. *American Journal of Political Science* 30, 666-687.

Krause, Volker and Suzuki, Susumu. 2005. Causes of Civil War in Asia and Sub-Saharan Africa: A Comparison. *Social Science Quarterly* 86:1, 160-177.

Lipset, Seymour Martin, 1959. Some Social Requisites of Democracy. Economic Development and Political Legitimacy. *American Political Science Review* 53:1, 69-105.

Marsden, Peter. 1998. *The Taliban: War, Religion and the New Order in Afghanistan*. Zed Books, London.

Marshall, Monty G. and Keith Jaggers 2007. *Polity IV Project: Dataset Users' Manual*. George Mason University and Center for Systemic Peace. http://www.systemicpeace.org/inscr/p4manualv2006.pdf

McLaughlin, Sara, Gates, Scott, Hegre, Håvard, Gissinger, Ranveig and Gleditsch, Nils Petter. 1998. Timing the Changes in Political Structures – A New Polity Database. *Journal of Conflict Resolution* 42:2, 231-242.

Munck, Gerardo L. and Verkuilen, Jay 2002. Conceptualizing and Measuring Democracy. *Comparative Political Studies* 35:1, 5-34.

Przeworski, Adam and Limongi, Fernando, 1997. Modernization: Theories and Facts. *World Politics* 49:2, 155-183.

Reilly, Benjamin, 2004. State Functioning and State Failure in the South Pacific. *Australian Journal of International Affairs* 58:4, 479-493.

Rubin, Donald B. 1976. Inference and Missing Data. *Biometrika* 63, 581-592.

Rubin, Donald B. 1996. Multiple Imputation after 18+ Years. Journal of the American statistical Association 91, 473-489.

Rubin, Barnett R., 2000. The Political Economy of War and Peace in Afghanistan. *World Development* 28:10, 1789-1803.

Sarkees, M.R., 2000. The Correlates of War Data on War. An Update to 1997. *Conflict Management and Peace Science* 18:1, 123-144.

Strand, Håvard. 2007. Measures and Mismeasures of Democracy. Working Paper. University of Oslo: Department of Political Science.

Vanhanen, Tatu, 2000. A New Dataset for Measuring Democracy, 1810–1998. *Journal of Peace Research* 37:2, 251–265.

Williams, Kieran. 1997. *The Prague Spring and its Aftermath: Czechoslovak Politics, 1968-1970*, Cambridge University Press, Cambridge.

Zack-Williams, Alfred B. 1999. Sierra Leone. The Political Economy of Civil War, 1991-1998. *Third World Quarterly* 20: 1, 143-162.

Table 4. Replication of Fearon and Laitin (2003) and robustness tests.

	model 4.1	model 4.2	model 4.3	model 4.4	model 4.5	model 4.6
	polity2	*polity2min*	*polity2pred*	standard mult. imp.	theory-based mult. imp.	mult. imp. comb. of 4.4 & 4.5
Prior war	-0.916	-0.870	-0.983	-0.905	-0.906	-0.907
	(0.312)***	(0.312)***	(0.384)**	(0.313)***	(0.312)***	(0.312)***
Per capita income	-0.318	-0.305	-0.334	-0.319	-0.318	-0.318
	(0.071)***	(0.071)***	(0.097)***	(0.071)***	(0.071)***	(0.071)***
log(population)	0.272	0.270	0.287	0.268	0.271	0.270
	(0.074)***	(0.074)***	(0.093)***	(0.074)***	(0.074)***	(0.074)***
log(% mountainous)	0.199	0.195	0.205	0.202	0.199	0.199
	(0.085)**	(0.085)**	(0.102)**	(0.085)**	(0.085)**	(0.085)**
Noncontiguous state	0.426	0.457	0.533	0.450	0.428	0.428
	(0.272)	(0.273)*	(0.422)	(0.272)	(0.272)	(0.272)
Oil exporter	0.751	0.728	0.287	0.760	0.750	0.752
	(0.278)***	(0.279)***	(0.365)	(0.279)***	(0.278)***	(0.278)***
New state	1.658	1.712	2.342	1.683	1.669	1.668
	(0.342)***	(0.342)***	(0.518)***	(0.342)***	(0.342)***	(0.342)***
Instability	0.513	0.570	0.536	0.564	0.529	0.529
	(0.242)**	(0.240)**	(0.321)*	(0.241)**	(0.242)**	(0.242)**
Ethnic fract.	0.164	0.160	0.554	0.168	0.163	0.163
	(0.368)	(0.369)	(0.513)	(0.369)	(0.369)	(0.369)
Religious fract.	0.326	0.301	-0.467	0.321	0.326	0.326
	(0.506)	(0.506)	(0.664)	(0.507)	(0.506)	(0.506)
Anocracy	0.521	0.373	0.514	0.396	0.488	0.489
	(0.237)**	(0.236)	(0.322)	(0.245)	(0.237)**	(0.237)**
Democracy	0.127	-0.111	-0.407	0.096	0.111	0.118
	(0.304)	(0.310)	(0.417)	(0.304)	(0.303)	(0.304)
Constant	-7.019	-6.922	-6.658	-6.952	-6.996	-6.997
	(0.751)***	(0.745)***	(0.953)***	(0.747)***	(0.749)***	(0.749)
Observations	6327	6327	3691	6327	6327	6327
Chi-sq/F test	1.73	2.37	4.73	.97	1.57	1.52
(H0: β(An.)=β(Dem.)	(p<.1881)	(p<.1240)	(p<.0296)	(p<.3253)	(p<.2098)	(p<.2169)

Notes: Standard errors in parentheses. * p < .10; ** p < .05; *** p < 0.01.

Table 5. Replication of Krause and Suzuki (2005) and robustness tests (Asia sample).

	model 5.1	model 5.2	model 5.3	model 5.4	model 5.5	model 5.6
	polity2	*polity2min*	*polity2 (det).*	standard mult. imp.	theory-based mult. imp.	mult. imp. comb. of 5.4 & 5.5
Per capita income	-1.744	-1.643	-1.554	-1.529	-1.519	-1.491
	(0.963)*	(0.895)*	(0.844)*	(0.884)*	(0.853)*	(0.839)*
log(population)	0.152	0.099	0.088	0.119	0.129	0.126
	(0.136)	(0.135)	(0.136)	(0.148)	(0.149)	(0.151)
log(% mountainous)	-0.492	-0.461	-0.418	-0.383	-0.372	-0.359
	(0.435)	(0.405)	(0.383)	(0.417)	(0.411)	(0.408)
Noncontiguous state	-1.946	-1.727	-1.612	-1.853	-1.881	-1.866
	(1.612)	(1.633)	(1.608)	(1.589)	(1.572)	(1.559)
Oil exporter	0.497	0.453	0.539	0.863	0.891	0.935
	(1.036)	(1.058)	(1.107)	(1.149)	(1.069)	(1.075)
New state	0.499	0.765	0.874	0.664	0.612	0.627
	(1.190)	(1.127)	(1.084)	(1.133)	(1.144)	(1.136)
Instability	0.334	0.498	0.511	0.494	0.482	0.492
	(0.510)	(0.506)	(0.520)	(0.541)	(0.515)	(0.518)
Ethnic fractionalization	4.202	4.349	4.150	3.700	3.572	3.494
	(2.136)**	(2.148)**	(2.050)**	(2.139)*	(2.172)*	(2.168)*
Religious fractionalization	2.932	2.495	2.240	2.464	2.497	2.434
	(2.785)	(2.529)	(2.358)	(2.673)	(2.675)	(2.646)
Militarization	0.585	0.560	0.545	0.566	0.567	0.563
	(0.295)**	(0.271)**	(0.258)**	(0.282)**	(0.281)**	(0.278)**
Trade Openness	-0.339	-0.298	-0.284	-0.326	-0.332	-0.331
	(0.144)**	(0.129)**	(0.122)**	(0.138)**	(0.139)**	(0.137)**
Regime	0.678	0.452	0.373	0.516	0.567	0.552
	(0.366)*	(0.258)*	(0.233)	(0.377)	(0.365)	(0.364)
Squared Regime	-0.031	-0.021	-0.018	-0.022	-0.024	-0.024
	(0.016)**	(0.011)*	(0.010)*	(0.017)	(0.016)	(0.016)
Constant	1.686	2.517	2.325	1.000	0.625	0.524
	(5.178)	(5.336)	(5.253)	(4.912)	(4.566)	(4.518)
Observations	609	609	609	609	609	609

Notes: Standard errors in parentheses. * p < .10; ** p < .05; *** p < 0.01.