

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/322223445>

# How (not) to measure democracy

Article · December 2017

---

CITATIONS

9

READS

6,999

1 author:



[Vanessa Alexandra Boese](#)

University of Gothenburg

25 PUBLICATIONS 377 CITATIONS

SEE PROFILE

# HOW (NOT) TO MEASURE DEMOCRACY

Vanessa A. Boese\*<sup>†</sup>

November 9, 2019

The final version of this paper is published at  
International Area Studies Review (June 2019)  
Available at: <https://doi.org/10.1177/2233865918815571>

## Abstract

Measures of democracy are regularly employed in the statistical analysis of economic, political and social policy. This paper reviews the measures' setup, strength and weaknesses across the three most prominent democracy datasets: the PolityIV, Freedom House and Varieties of Democracy (V-Dem) data. The measures developed by the V-Dem project outperform Polity2 and Freedom House Index with respect to the underlying definition, measurement scale as well as the theoretical justification of the aggregation procedure. The three indices display a high level of agreement for those observations included in all three datasets. The most substantial differences between the indices lie in the indices' coverage, i.e. in their non-missing observations (in Polity2 coding, for example, years during which a country is occupied by foreign powers constitute missing values), the availability of disaggregate data and the above mentioned key areas. This paper clarifies when to proceed with caution but for the most part advocates the use of V-Dem in the statistical analysis of democracy.

**Keywords:** democracy, Polity, Freedom House, Varieties of Democracy (V-Dem), measurement of democracy

## 1 Introduction

The characteristics of a country's system of governance are essential for the understanding of basic questions in political science and economics.<sup>1</sup> Over the past thirty years systems of governance have been measured in several different ways. Numerous studies have demonstrated that the results vary with the index used.<sup>2</sup> Hence, it is of utmost importance to understand the specific setup of the existing indices before deciding which index to use. This paper facilitates the decision which index to use when by examining the three most commonly used democracy datasets: The PolityIV, Freedom House and Varieties of Democracy (V-Dem) data. The measures developed by the V-Dem project outperform Polity2 and Freedom House Index (FHI) with respect to the underlying definition, measurement scale as well as the theoretical justification of the aggregation procedure. This article illustrates under which circumstances the three democracy measures code countries dis-/similarly and points out those frameworks under which certain measures can/should not be employed.

---

\*Email: boesevan@hu-berlin.de, School of Business and Economics, Humboldt-Universität zu Berlin, Germany.

<sup>†</sup>I would like to thank two anonymous reviewers, Scott Gates, Carl Henrik Knutsen, Håvard Strand, Katharina Lehmann-Uchner, Steffen Elstner, Karim Diebold, and the participants of the Oxford Development Economics Workshop 2017 as well as those of the DENeB discussion round 2017 for excellent comments and suggestions. All errors are my own.

<sup>1</sup>Hence, they have been frequently examined, for example by Acemoglu et al. (2014), Franco et al. (2004), Przeworski et al. (2000).

<sup>2</sup>For example Bogaards (2010); Bollen (1980); Bollen & Jackman (1985), and Bollen & Jackman (1989); Casper & Tufis (2003); Cheibub et al. (2010); Elkins (2000); Treier & Jackman (2008).

The question of how to correctly quantify a state's system of governance for statistical analyses has been a highly contested subject in applied political and economic research.<sup>3</sup> The discussion in today's literature evolves around three central questions:<sup>4</sup>

1. Definition: along which components should a system of governance be defined?
2. Measurement: how to quantify these components and
3. Aggregation: how to combine them into a single index of democracy?

There is no consensus in current literature on any of these questions. Due to the multifaceted nature of the concepts "system of governance" or "democracy" a one-size-fits-all answer to the first question is unlikely to exist.<sup>5</sup> The characteristics of a system of governance relevant for a study vary with the research question at hand. This has led to the creation of numerous distinct democracy indices.<sup>6</sup> Most of the democracy indices used in the empirical literature rely on different definitions of democracy. They capture different aspects of a state's authority and decision finding structure. Several papers exist examining the strength and weaknesses of democracy measures.<sup>7</sup>

The suitability<sup>8</sup> of a democracy measure for a given statistical application depends on the research framework. However, its internal validity and reliability can be examined independently. The question "does the democracy measure capture what it is supposed to?" is generally referred to as the concept of validity. Note, that the "what it is supposed to measure"-part varies with the underlying definition of each democracy measure. "How well/ how precise does the democracy index measure what it is supposed to?" on the other hand refers to an index's reliability.<sup>9</sup> The two concepts are, by definition, highly interlaced. So how to assess and compare the validity and the reliability of existing democracy measures? Both, validity and reliability, depend on the underlying definition and measurement level, the data collection and aggregation procedure. Consequently, these are the central quality criteria this paper will examine for Polity2, FHI as well as for the V-Dem indices, in particular for their Electoral Democracy Index, Polyarchy. Evaluating its implicit assumptions and definitions as well as their implications for statistical analyses is crucial for choosing an appropriate democracy measure. To understand the conceptual differences underlying the definition and setup of the democracy measures each of the three central questions stated above will be briefly summarized in the remainder of this section.

## 1.1 Definition of democracy

In current debates on measures quantifying of systems of governance ("democracy indices") it has become the norm to use the term 'democracy' pars pro toto, i.e. synonymously to "system of governance". This paper follows this convention. Nevertheless, from a theoretical point of view it is important to acknowledge that the democracy indices are in fact measuring observed levels of authority patterns or systems of governance. The word 'observed' is key here: following Dahl (1971), Goertz (2006) and Treier & Jackman (2008) this paper distinguishes between the concept of democracy as a theoretical construct and its actual observable manifestations, i.e. democracy is viewed as a latent,

---

<sup>3</sup>Boix et al. (2013), give a detailed overview of the current debates on measuring democracy, Table 1, pp.1526. Please refer to their paper for a more detailed listing of all available comments to the debate. This paper focuses on the contributions concerning Polity2, Freedom House and the V-Dem indices.

<sup>4</sup>Compare Munck & Verkuilen (2002), Table 2, p. 8; Boix et al. (2013), pp. 1525; Coppedge et al. (2011), p.248.

<sup>5</sup>Crick (2002), p.1, for example argues that the term "democracy" might be impossible to define "because the very definition carries a different social, moral or political agenda."

<sup>6</sup>Munck & Verkuilen (2002), Table 1, p. 6, and Table 3, p.10, present a detailed lists of available indices.

<sup>7</sup>For example Cheibub et al. (2010); Coppedge et al. (2008); Coppedge et al. (2011); Coppedge, Gerring, Lindberg, Skaaning & Teorell (2017); Elkins (2000); Gleditsch & Ward (1997); Höglström (2013); Knutsen (2010); Munck & Verkuilen (2002); Munck (2009); or Treier & Jackman (2008).

<sup>8</sup>Suitability here refers to whether the measurement level and the democracy definition are appropriate for the given analysis.

<sup>9</sup>For a discussion of validity and reliability see, for example, Coppedge, Gerring, Lindberg, Skaaning & Teorell (2017), pp.16; and Munck (2009), pp. 23.

continuous quantity. A democracy measure's validity therefore can be thought of in terms of concept-measure consistency. To obtain a valid measure of democracy, the well-defined concept should drive the methodology.<sup>10</sup> A consequence of this latency is the uncertainty/noise in coding of the empirical measures.<sup>11</sup> In addition to noise in the measures coding due to the latency of democracy several studies, such as Bollen (1993), Bollen & Paxton (2000) or Elff & Ziaja (2018), show that method factors (i.e. coder-specific systematic biases in each dataset) can provide further sources of uncertainty in democracy measures. To avoid false inference a precise definition of the concept of democracy<sup>12</sup> guiding the methodological construction of the empirical index is essential. Following Goertz (2006) and Dahl (1971) perfect democracy and autocracy can be thought of as the poles between which a continuum of polyarchies/authority patterns/systems of governance unfolds.

Even on a conceptual level the question of how to conceptualize a system of governance is an entire field of research.<sup>13</sup> Schumpeter's minimalist definition<sup>14</sup> captures the core trait of an institutional definition of democracy: **a regime holding contested elections - contestation and participation.**<sup>15</sup> It is worth noting, that these two democratic principles are complements: a country cannot be considered democratic if there is full participation, but no contestation and vice versa.

The central criticism of this minimalist institutional definition is that by acknowledging only the electoral dimension other defining components of democracy are ignored.<sup>16</sup> Instead of defining democracy institutionally Dahl (1971) goes back to the word's Greek origin.<sup>17</sup> He takes a substantive approach and views **popular control over collective decision making as well as political equality** as core elements of democracy.<sup>18</sup> He recognizes that to realize this democratic principle of popular rule and political equality a number of institutional guarantees have to be met. These institutional guarantees are a necessary (not sufficient!) condition to ensure the implementation of the two core institutional democracy traits of contestation and participation. In that sense there cannot be a democratic regime without holding contested elections but at the same time not all regimes holding contested elections classify as democracies.<sup>19</sup>

On a conceptual level, a distinction between *de facto* and *de jure* attributes of a polity can be made. This paper regards the formal institutional setup of a polity as its *de jure* framework. Observable policy outcomes on the other hand constitute *de facto* elements. This concept of *de facto* encompasses the way the institutions operate in practice as well as the performance and practices generated by the actors in the system.<sup>20</sup> Political violence carried out by non-governmental actors for example can be an byproduct of the "institutional output" unintended by the system's setup and consequently be part of the *de facto* reality.

---

<sup>10</sup>See Goertz (2006), p.11.

<sup>11</sup>Treier & Jackman (2008) show this for the Polity index and Høyland et al. (2012) show it for index rankings using (among others) the FHI.

<sup>12</sup>The point cannot be stressed enough that this precise definition of the concept might vary with application.

<sup>13</sup>See for example Schmitter & Karl (1991); Crick (2002); Huntington (1993), pp.5; Diamond (1999), pp. 7; or Dahl (1971), chapter 1.

<sup>14</sup>"(...) the democratic method is that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote" Schumpeter (1976), p.269.

<sup>15</sup>Participation here is used in an inclusiveness/suffrage sense.

<sup>16</sup>See, for example, Diamond (1999), p.9; Mainwaring et al. (2007), p.128; Schmitter & Karl (1991), p. 78.

<sup>17</sup>"δημοκρατία" is a compound of δῆμος - the common people, citizens and κράτος - rule; thus literally meaning "rule of the people".

<sup>18</sup>See Dahl (1971), p.1.

<sup>19</sup>See Cheibub et al. (2010), p. 72.

<sup>20</sup>An anonymous referee's comments were very helpful in substantiating this definition and are very much appreciated.

Lastly, it is not enough to identify components to include in a definition of democracy. The theoretical relations of these components with each other need to be taken into account. Only once a logical structure between these elements is established they can be aggregated into a measure of democracy in a meaningful way.<sup>21</sup> Not ordering the elements vertically by their level of abstraction gives rise to problems of redundancy and conflation (joining elements which are symptoms of distinct overarching attributes).<sup>22</sup>

## 1.2 Measurement Scale of the Democracy Measure

The question how to quantify the components and on what kind of scale to map the empirical distribution of cases should also be determined by the concept behind the measure. Most polychotomous democracy indices, such as Polity2 or FHI, are coded on a nominal or ordinal scale. However, in empirical research these indices are often treated as if they were coded on an interval scale even though there is no numerical relationship between the categories. Doing so implies that the differences between the values have a meaningful interpretation. This is a major point of criticism of several of the polychotomous indices.<sup>23</sup> While the definition and aggregation procedure of the V-Dem democracy indices<sup>24</sup> permits using them as quasi continuous measures, this is not the case for Polity2 and FHI. Polity2 is defined on a scale which is not even ordinal,<sup>25</sup> and differences between categories have no meaningful interpretation. One specific FHI score can be the result of a variety of different underlying factors. The Freedom House aggregation procedure using equal weighing and collapsing further contributes to its final measurement scale at best being ordinal.

Concept-measure consistency is also essential in terms of the scale's end points. Whether a democracy index can take on top- or bottom values should be determined by whether the concept of democracy allows for perfect democracy and autocracy endpoints. Assuming a continuous concept the ends of the conceptual scale can be thought of as unreachable poles (one can get infinitely close to but never reach them). Fixed endpoints imply that one believes in a state of perfect autocracy/democracy. From the empirical point of view there is a disadvantage to attainable top- and bottom values which becomes obvious with long time series. For example, the increased number of democracies in the system of states after the end of the Cold War made it impossible to examine "the societal dynamics associated with the consolidation and maintenance of democratic authority patterns"<sup>26</sup> within these states using the Polity Coding Scheme.

## 1.3 Collection, Aggregation and Weighing of Components

The question of how to adequately collect, weigh and aggregate components has been subject to much debate.<sup>27</sup> The most important discussion points are displayed in Figure 1. The validity and reliability of a measure increase the more adequately these points are addressed.

The first principle of aggregation acknowledges the fact that the type of information captured in the disaggregated data is essential.<sup>28</sup> Several democracy measures are based on components coded using

---

<sup>21</sup>Munck & Verkuilen (2002) p. 13, provide an overview of such a logical organization of elements.

<sup>22</sup>Munck & Verkuilen (2002), for example, point out that the Polity IV Data's aggregation rule suffers from a redundancy problem, p.13.

<sup>23</sup>See for example Gleditsch & Ward (1997), p. 380, point 1; Cheibub et al. (2010), p. 75 (FHI), p. 76 (Polity Index); Boix et al. (2013), p. 1529.

<sup>24</sup>They are aggregated using a mixture of Bayesian item response theory measurement models, addition and multiplication (see section *Varieties of Democracy (V-Dem) Dataset*).

<sup>25</sup>See section *Polity Index*. Cases of interregnum/anarchy, for example, are coded as 0 - the "middle" of perfect democracy and perfect autocracy.

<sup>26</sup>See Marshall et al. (2017a), p.10.

<sup>27</sup>See for example Cheibub et al. (2010), pp 74; Coppedge et al. (2011), pp. 250; Boix et al. (2013) p. 1530; Knutsen (2011), pp. 83; Munck & Verkuilen (2002), pp. 22.

<sup>28</sup>Cf. Cheibub et al. (2010), p. 74.

Table 1: The principles of aggregation

1. Data Collection Process: type of information used
  - clear and transparent rules for the collection of information
  - public availability of disaggregate data
  - sources: subjective evaluations vs. objective facts
2. Aggregation Rule: how information is aggregated and weighted
  - choice of level of aggregation
  - capture of underlying theory (reflecting an accepted definition of democracy) in the aggregation rule
  - Provision and justification of the aggregation rule
    - substitutability vs. complementary relations
    - factoranalysis or principal component analysis
    - addition, multiplication
    - openness to further tests

subjective inferences "and perhaps even guesses", Cheibub et al. (2010), p. 77. Including of subjective evaluations reduces replicability and increases measurement error, but due to democracy being a social construct a limitation to formal (and better quantifiable) indicators leads to lower validity.<sup>29</sup> Replicability is a key feature for a useful democracy measure. Hence, clear and transparent rules for the information collection process and public availability of disaggregate data should be considered crucial.<sup>30</sup> This point has not been adequately addressed by many of the existing indices, e.g. the FHI, and consequently has been frequently criticised.<sup>31</sup>

The second principle of aggregation concerns the Aggregation Rule, i.e. how the disaggregate information is weighted and aggregated. First, a suitable level of aggregation should be determined.<sup>32</sup> While a high level of aggregation (aggregating all components and indicators into one democracy index) is appealing for the use of a democracy index in empirical calculations it also entails loss of complexity, information and validity.<sup>33</sup> Second, the underlying theory (reflecting an accepted definition of democracy)<sup>34</sup> should be captured in the aggregation rule.<sup>35</sup> Third, to increase replicability the aggregation rule should be provided and justified.<sup>36</sup> This entails for example a justification of choice of relation between the components: are they substitutes or complements? A non-arbitrary aggregation rule justifies its use of addition or multiplication of components or even the use of factor or principal component analyses. And, last but not least, it is open to further examinations.<sup>37</sup>

Democracy as a political concept is inherently difficult to define and to capture in a quantitative way. Yet, this paper demonstrates that the V-Dem project succeeded in constructing quantitative measures excelling in the key areas mentioned above. The goal of this paper is thus to provide a thorough overview of relevant issues to consider when choosing a democracy measure for a statistical application. It seeks to give the reader an understanding of the differences, weaknesses and strengths of the three democracy datasets.

---

<sup>29</sup>Cf. Knutsen (2011), p. 84.

<sup>30</sup>Cf. Cheibub et al. (2010), p. 74; Munck & Verkuilen (2002), Munck & Verkuilen (2002), p. 25 & 26.

<sup>31</sup>For example by Cheibub et al. (2010), pp. 75, or Munck & Verkuilen (2002), Munck & Verkuilen (2002), p. 25.

<sup>32</sup>Cf. Munck & Verkuilen (2002), p. 23.

<sup>33</sup>Cf. Munck & Verkuilen (2002), p. 22.

<sup>34</sup>Cf. Coppedge et al. (2011), p. 250.

<sup>35</sup>Cf. Knutsen (2011) p. 83.

<sup>36</sup>See also Munck & Verkuilen (2002).

<sup>37</sup>Cf. Munck & Verkuilen (2002), p. 25.

Table 2: Overview of the democracy measures analyzed in this paper

Index	Definition includes	Availability from to	N	Range	Aggregation rule	Advantages	Disadvantages
Freedom House Index	Subjective lists of civil liberties and political rights	1972 2016	209	1 (highest) to 7 (lowest)	Addition, ranking (including collapsing variation) and averaging	- Covers a variety of democratic features (civil liberties and political rights)	- Maximalist (subjective) democracy definition - Definition changed over time - Aggregation rule without justification - Measurement problems - Disaggregate data not publicly available - Unclear distinction between de facto and de jure aspects of a polity and their influence in the democracy measures
Polity2 Index	Contestation of offices, constitutional constraints on head of state	1800 2016	195	-10 (lowest) to 10 (highest)	Combination of weighing and addition	- Broad temporal & spatial scope - Detailed coding rules - Disaggregate data publicly available	- No theory behind aggregation rule - Definition: omission of suffrage/any participation - Factionalism categorization - 0 coding for interregnum - Missing values for foreign interruption
V-Dem Democracy Indices	Electoral, liberal, egalitarian, deliberative, and participatory dimension	1900 2016	177	0 (lowest) to 1 (highest)	Mixture of Bayesian item response theory measurement models, addition and multiplication	- Broad temporal & spatial scope - Weakest link argument included in aggregation procedure - Theoretical justification of aggregation rules - Bridge- and Lateral-coding - Disaggregate data publicly available - Public discussion of measurement error - Ordinal versions of all variables offered in addition	- Unclear distinction between de facto and de jure aspects of a polity and their influence in the democracy measures

$N$  = number of countries/territories coded. Range: (lowest/highest) refers to the lowest/highest possible level of democracy

Table 3: Components of the Polity Index and the democratic concepts they capture

Competitiveness of Executive Recruitment	XRCOMP	contestation
Openness of Executive Recruitment	XROPEN	contestation
Constraint on Chief Executive	XCONST	institutional constraints
Competitiveness of Political Participation	PARCOMP	contestation
Regulation of participation	PARREG	contestation

## 2 Democracy Indices

The number of existing democracy indices is too vast to give a detailed overview in one paper.<sup>38</sup> Some indices, such as the V-Dem indices are available in country-event format. For the sake of increased comparability among indices and the usability in classic time-series cross-country studies this paper focuses on data in country-year format. Table 2 summarizes the three main democracy measures discussed in this section. It displays the measurement scale, the democracy definition, the measure’s temporal and geographical scope, its’ range, aggregation rule as well as its’ strengths and weaknesses.

### 2.1 Polity Index

Due to its broad chronological (1800 - 2016) and geographical scope (195 countries) the Polity Index is one of the most frequently used democracy indices in current research. When it was first introduced in 1975 it constituted one of the first major attempts to quantify authority patterns on a global scale over an extended period of time. These first codings were based on the detailed theories of authority patterns put forward by Eckstein (1973) and Gurr (1974). The first Polity data was coded with the explicit objective of answering the question whether ”the durability (persistence and adaptability) of political systems depends at all upon the nature of their structures of political authority”.<sup>39</sup> Said structures of political authority, i.e. the institutional framework therefore form the core of the Polity Index.

Today the data is assembled by researchers from the Polity IV Project<sup>40</sup> at the Center for Systemic Peace. The most recent version of data available is the Polity IV Dataset Version 2016, Marshall et al. (2017b).<sup>41</sup>

The Polity Index ranges from -10 (strongly autocratic) to 10 (strongly democratic) and is calculated by subtracting a measure of autocracy (Autoc) from a measure of institutionalized democracy (Democ):  $Polity = Democ - Autoc$ . Both indices, Democ and Autoc, range from 0 to 10. They are made up of scores of five different components reflecting the polity’s executive recruitment (XRCOMP<sup>42</sup> and XROPEN<sup>43</sup>), its’ constitutional constraints (XCONST<sup>44</sup>) and its’ political participation (PARCOMP<sup>45</sup> and PARREG<sup>46</sup>) as presented in Figure 3.

<sup>38</sup>For a quick overview please refer to Coppedge, Gerring, Lindberg, Skaaning & Teorell (2017), Table 1; Munck (2009), Table 4.2 or Pemstein et al. (2010), Table 1.

<sup>39</sup>See Gurr (1974), p.1482.

<sup>40</sup>More information on the PolityIV Project is available at: <http://www.systemicpeace.org/polityproject.html>.

<sup>41</sup>Two types of datasets are offered: Polity IV Dataset in country-year format and the Polity IVd Dataset, where ”d” denotes the country-date format.

<sup>42</sup>”Competitiveness refers to the extent that prevailing modes of advancement give subordinates equal opportunities to become superordinates”, Marshall et al. (2017a), p. 21.

<sup>43</sup>”Recruitment of the chief executive is ”open” to the extent that all the politically active population has an opportunity, in principle, to attain the position through a regularized process”, Marshall et al. (2017a), p. 22.

<sup>44</sup>”(…) the extent of institutionalized constraints on the decisionmaking powers of chief executives, whether individuals or collectivities. Such limitations may be imposed by any ”accountability groups.” (….) The concern is therefore with the checks and balances between the various parts of the decision-making process.”, Marshall et al. (2017a), p. 24.

<sup>45</sup>”(…) the extent to which alternative preferences for policy and leadership can be pursued in the political arena.”, Marshall et al. (2017a), p. 26.

<sup>46</sup>”Participation is regulated to the extent that there are binding rules on when, whether, and how political preferences are expressed.”, Marshall et al. (2017a), p. 25.



Table 4: Aggregation Rule of the Polity Index. Source: Polity IV Dataset Users' Manual, Marshall et al. (2017a), p.15, 16 and 26

$$A = Autoc, A < 0; D = Democ$$

$$Polity = Democ + Autoc$$

Authority Coding:	Scale Weight	Counted in...
<u>Competitiveness of Executive Recruitment (XRCOMP):</u>		
(1) Selection	-2	A
(2) Transitional	+1	D
(3) Election	+2	D
<u>Openness of Executive Recruitment (XROPEN):</u>		
for DEMOC: only coded if XRCOMP is Election (3) or Transitional (2)		
for AUTOC: only coded if XRCOMP is coded Selection (1)		
(1) Closed	-1	A
(2) Dual/designation	-1	A
(3) Dual/election	+1	D
(4) Election	+1	D
<u>Constraint on Chief Executive (XCONST):</u>		
(1) Unlimited authority	-3	A
(2) Intermediate category	-2	A
(3) Slight to moderate limitations	-1	A
(4) Intermediate category	+1	D
(5) Substantial limitations	+2	D
(6) Intermediate category	+3	D
(7) Executive parity or subordination	+4	D
<u>Competitiveness of Political Participation (PARCOMP):</u>		
(0) Not Applicable		not counted in democ/autoc
(1) Repressed	-2	A
(2) Suppressed	-1	A
(3) Factional	+1	D
(4) Transitional	+2	D
(5) Competitive	+3	D
<u>Regulation of participation (PARREG):</u>		
AUTOC ONLY		
(1) Unregulated		not counted in democ/autoc
(2) Multiple Identity		not counted in democ/autoc
(3) Sectarian	-1	A
(4) Restricted	-2	A
(5) Regulated		not counted in democ/autoc

Note that these components capture the people's participation only with respect to participation in the political process, but no suffrage requirement is included. Hence "participation" as used by the Polity Project is not equal to the use of the term in the debates on measuring democracy. Rather the participation components evaluate the extent to which oppositional political activity is possible and regulated. The Polity Index is in fact a measure of political contestation rather than democracy - even if one embraces the minimalist democracy definition with contestation and participation.<sup>47</sup>

Figure 4 depicts the aggregation rule of Polity and the weighing scheme for Democ and Autoc. A country-year receives a score (scale weight) for each component. The points scored for Autoc are subtracted from the Democ index to obtain the Polity Index. Aside from a vague theory (explaining the "logic of institutionalized democracy and autocracy", Marshall et al. (2017a), p. 15 and p. 16), there is no justification given for the weighing and aggregation rule. Each of the component variables (XRCOMP, XROPEN, XCONST, PARCOMP, PARREG) is coded using three or more categories.

<sup>47</sup>In the Online Appendix a factor analysis of the Polity Index' components is presented. It shows that the variation in the components can be explained by one latent factor - political contestation.

However, not all of the categories are taken into account when calculating the democ/autoc and consequently the Polity Index. In a thorough examination of the Polity aggregation procedure Goertz (2006) shows that the Polity Index suffers from concept-measure inconsistency, i.e. the measure does not capture what the underlying concept of democracy very well.

Since the idea behind Polity was to quantify institutional frameworks it reached the limits of its domain in cases where there was no regular institutional setup. Years in which a country's central political authority is collapsed are considered as an interregnum period and coded "-77" on Democ, Autoc and Polity. Years in which central authority is taken over by foreign powers are considered interruption periods and coded "-66" for each index. During transition periods in which "new institutions are planned, legally constituted, and put into effect"<sup>48</sup> the indices receive a "-88" coding. Of 17,228 observations in the Dataset Version 2016, this is the case for 772 observations, i.e. around 4.5% of the observations. These three categories limit the use of the Polity Index in empirical research. To integrate these cases into the Polity scheme and make them accessible for quantitative studies the Polity IV Project introduced Polity2. It is defined as follows:

$$Polity2 := \begin{cases} \text{Polity Index,} & \text{if Polity Index} \in [-10, 10] \\ \text{missing value,} & \text{if Polity Index} = -66 \text{ (foreign interruption)} \\ 0, & \text{if Polity Index} = -77 \text{ (interregnum)} \\ \text{prorated across transition,} & \text{if Polity Index} = -88 \text{ (transition)} \end{cases}$$

Even though this definition enables the inclusion of transition and interregnum cases in time series and cross section models it creates a range of problems: First, the foreign interruption observations are still missing. In the Dataset Version 2016 there are 233 such cases, i.e. 1.3% of all observations. Depending on the goals of the research at hand, especially when examining democratic transitions/autocratic backsliding or democracy and civil conflict, these missing cases are of particular interest. Second, a Polity2 value of "0" can occur in three different cases.<sup>49</sup> The most intuitive one is if a country's Autoc score equals its Democ score as it does for example in the case of Albania in 1996. The second reason why a country might be assigned a Polity2 value of "0" in a given year is a transition period. If a country is undergoing such a transition its Polity2 value is prorated across the time span of the transition. For example, if it has a transition year in 1990 and index values of "-2" in 1989 and "2" in 1991, then 1990 will be assigned a "0". Both of these codings are still somewhat suitable for empirical research since the differences between the Polity2 values remain meaningful. However, this is not true for the third case in which a country might be assigned a value of "0". In years in which a total collapse of central political authority occurs the country is assigned a Polity2 value of "0". The civil war in Afghanistan, 1992 - 1995, is an example for such a case of "anarchy". The meaning behind this particular "0" is rather different from the others and it renders differences between Polity2 values impossible to interpret. This "0-coding" hence affects the measurement scale: Polity2 is coded on a nominal (not even an ordinal!) scale limiting the usability of Polity2 in econometric models (at least if the "0-coding" is maintained).

In the Polity III Data (a predecessor of the Polity IV Dataset) some categories of the components on competition (PARCOMP) and regulation of political participation (PARREG) were defined with explicit reference to conflict.<sup>50</sup> Hence, countries experiencing severe civil conflict were highly unlikely to be classified as high/low democracies. Rather they were categorized as semi-democracies/anocracies leading to a number of studies<sup>51</sup> examining why semi-democracies seemed to be more prone to conflict than "pure" democracies/autocracies. Being aware of this point of criticism the Polity IV Project

<sup>48</sup>Marshall et al. (2017a), p. 19.

<sup>49</sup>Gleditsch and Ward criticised a similar point, namely that each Polity Index value can be achieved by a large number of different scores in the components and thus reflecting quite different polities in Gleditsch & Ward (1997).

<sup>50</sup>PARCOMP, categories (0 - unregulated) and (1 - factional) were coded with reference to civil war and violent conflict in the Polity III Data. Similarly, PARREG category (factional/restricted) exhibited the same problem, see Vreeland (2008), p. 406.

<sup>51</sup>Some of the most notable ones are Hegre (2001); Vreeland (2008); Fearon & Laitin (2003), pp. 84; Goldstone et al. (2010).

removed the explicit references from the definitions of the components.<sup>52</sup> However, even though the categories do not include the explicit reference anymore, they still capture the concept of a country undergoing conflict.<sup>53</sup> Due to the categorization of interregnum, interruption and transitions as described above and this coding problematic the aggregate Polity2 should not be used in research on civil conflict.

One of the strongest advantages of the Polity IV Dataset is the availability of the disaggregate data.<sup>54</sup> This enables the breakdown of Polity2 into its components. Once the issues mentioned above (factionalism, missing interruption values etc.) are adequately addressed the components can be reassembled to form a measure of political contestation that (while still not continuous) can be employed in certain empirical models.

## 2.2 Freedom House

Freedom House is a US-based non-governmental organization dedicated to the promotion of freedom and democracy worldwide. Founded in 1941 in the midst of WWII it has significantly amplified its sphere of operation and influence with the publication of annual "Freedom in the World" reports since 1973. The reports are based on annual surveys of global political rights and civil liberties. The survey data is available for 209 countries and territories from 1972 to 2016. With 209 countries covered FHI is the most inclusive of the three measures studied. The first reports and ratings were single-authored by Raymond Gastil, who argued in Inkeles (1991): "By working alone the author has not had to integrate the judgments of a variety of people. The hunches and impressions that are so important in a survey of this kind would be almost impossible to keep on the same wave lengths if one had an Asianist, Africanist, and Latin Americanist to satisfy before the ratings were finalized for each year".<sup>55</sup> While today the reports are produced by a team of "external analysts"<sup>56</sup>, the checklist question framework introduced by Gastil and its inherent subjectivity remains. The FHI is not built upon any theoretical concept of democracy or freedom, rather it is a country comparison of an undefined concept of "freedom" based upon said "hunches and impressions". Bush (2017) shows that these impressions correspond to a large degree to the US elite's perception of other countries systems of governance and hence proposes taking FHI as a measure of such.<sup>57</sup> The remainder of this section will further illustrate this by providing a short overview of the components and aggregation procedure as well as the most prominent points of criticism.<sup>58</sup>

The Freedom in the World Index (FHI) evaluates the freedom concept along two dimensions: freedom of political rights and civil liberties. It is assembled in three steps. First, the Freedom House coders award from 0 (smallest) to 4 (greatest degree of freedom) points/scores to 27 questions. 12 of these questions regard the political rights dimension while the remaining 15 questions address the implementation of civil liberties.<sup>59</sup> Instead of four clear cut possible answer categories (0 - 4 points are rewarded) for each of these questions a number of subquestions are given to clarify the concept. The disaggregate data is not publicly available, hence there is no way of knowing or replicating how a country achieved a certain score or of testing the implications of the aggregation rule.<sup>60</sup>

---

<sup>52</sup>See Marshall et al. (2017a), pp. 25.

<sup>53</sup>See for example the definition of PARCOMP's category (3 - factional competition): "Polities with parochial or ethnic-based political factions that regularly compete for political influence in order to promote particularist agendas and favor group members to the detriment of common, secular, or cross-cutting agendas.", Marshall et al. (2017a), p. 27.

<sup>54</sup>Also noted by Munck & Verkuilen (2002), p.20.

<sup>55</sup>See Inkeles (1991), p. 22.

<sup>56</sup>See House (2017a).

<sup>57</sup>See Bush (2017), p.725.

<sup>58</sup>Since these points of criticism are plenty but the number of pages is not, the critiques are discussed briefly and the interested reader is asked to refer to the corresponding articles for further information.

<sup>59</sup>The Online Appendix provides a table displaying the concrete questions and respective scores.

<sup>60</sup>This has been frequently criticised, for example by Munck & Verkuilen (2002), p. 25; Munck (2009), p. 29; Cheibub et al. (2010), p. 75.

Table 5: Score to rating conversion for the Political Rights Index, Source: House (2017b)

Total Scores		Political Rights Rating	
36	- 40	1	Greatest range of political rights implemented
30	- 35	2	Intermediate category (between 1 and 3)
24	- 29	3	Countries and territories with a rating of 3, 4, or 5 either moderately protect
18	- 23	4	almost all political rights or strongly protect some political rights while
12	- 17	5	neglecting others. The same factors that undermine freedom in countries with a rating of 2 may also weaken political rights in those with a rating of 3, 4, or 5, but to a greater extent at each successive rating.
6	- 11	6	Intermediate category (between 5 and 7)
0	- 5	7	Few or no political rights (...) sometimes in combination with civil war

In a second step, depending on the sum of scores obtained in the political rights and civil liberties components a rating is assigned. The scores to rating conversion for both components is presented in Table 5 for the Political Rights Index.<sup>61</sup> The conversion for the Civil Liberties is carried out analogously. It is worth noting, that (due to the high number of possible ways of obtaining a certain score rating) each of the seven categories captures numerous countries with very distinct political rights and civil liberties.

In a third step the political rights and civil liberties ratings are averaged to form the freedom rating. It ranges between 1 and 7. This rating is used to categorize the countries into three groups: Free (rating between 1 - 2.5), Partly Free (rating between 3 - 5) and Not Free (rating between 5 - 7). As mentioned above, each of the seven categories contains a wide range of countries with politically very different environments. Further categorizing them into three groups inherently continues disguising this variation.<sup>62</sup>

One of the most criticized aspects of the Freedom House data is the compilation of components by means of checklist questions without a theoretical justification. This is problematic in several ways: First, the components are not ordered by level of abstraction and the relationship between the components is not considered. This gives rise to the problem of conflation<sup>63</sup> as criticized by Munck & Verkuilen (2002), p. 14.; Coppedge et al. (2011), rightfully observe that "the high inter-correlations of the Freedom House indicators coupled with their ambiguous coding procedures suggest that these components may not be entirely independent of one another".<sup>64</sup> Second, as mentioned above there are no clear cut answers for the checklist questions. Since the questions are formulated in a way as to capture highly subjective features (e.g. "Are the electoral laws and framework fair?") the lack of clear answers transmits this problem of subjectivity further into the data. This was, for example, criticized by Cheibub et al. (2010), p. 75.

<sup>61</sup>Note, that while Table 5 depicts 0 as the minimum score a country can be given "It is possible for a country's or territory's total political rights score to be less than zero (between -1 and -4) if it receives mostly or all zeros for each of the 10 political rights questions and it receives a sufficiently negative score for political rights discretionary question B. In such a case, it would still receive a final political rights rating of 7", House (2017b). The discretionary political rights questions (see Online Appendix) reduce the political rights scores without a clear rule of application thus contributing to a further subjective bias in the data.

<sup>62</sup>"For example, those at the lowest end of the Free category (2 in political rights and 3 in civil liberties, or 3 in political rights and 2 in civil liberties) differ from those at the upper end of the Free group (1 for both political rights and civil liberties). Also, a designation of Free does not mean that a country or territory enjoys perfect freedom or lacks serious problems, only that it enjoys comparatively more freedom than those rated Partly Free or Not Free (and some others rated Free)", House (2017b).

<sup>63</sup>See section *Definition of democracy*.

<sup>64</sup>See Coppedge et al. (2011), p. 251.

Another point of concern is the inappropriate aggregation rule of addition - equal weighing. Assigning equal weights to each question asked/concept contained is disputable in light of their content. This aggregation rule does not capture the complementarity of the concepts participation and contestation. Furthermore, it does not view them necessary conditions for a democracy.<sup>65</sup> It is also noteworthy that the measurement scale is neither continuous nor ordinal, it is at best categorical.<sup>66</sup>

There is also concern with respect to the usage of the FHI in a time series context. A number of studies have exposed an ideological bias in the Freedom House Data: Bollen & Paxton (2000), p. 77, for example find evidence for a systematic downrating of Marxist-Leninist countries, especially in Freedom House's early years and abating around the mid 1980s. In addition, in some years the coding rules are altered from one year to the next and previous years are not updated. Cheibub et al. (2010), p. 75, and Munck (2009), p. 148 (footnote 15) observe that hence, the use of Freedom House Data in a time series context is hardly justifiable.

The problem of subjectivity is also inherent in the coding of the time series. Freedom House states: "the scores from the previous edition are used as a benchmark for the current year under review. A score is typically changed only if there has been a real-world development during the year that warrants a decline or improvement (e.g., a crackdown on the media, the country's first free and fair elections), though gradual changes in conditions, in the absence of a signal event, are occasionally registered in the scores".<sup>67</sup> This benchmarking can potentially lead to the transmission of a subjective coding bias over long periods of time.

In conclusion, Freedom House Data should be used in statistical analyses with extreme caution. The Dataset "by Freedom House (...) exemplifys problems in all three areas of conceptualization, measurement, and aggregation", Munck & Verkuilen (2002), p. 28.

### 2.3 Varieties of Democracy (V-Dem) Dataset

The Varieties of Democracy (V-Dem) Dataset is assembled by a cooperation of over 50 scholars from all over the world, co-hosted by the Department of Political Science at the University of Gothenburg, Sweden, and the Kellogg Institute at the University of Notre Dame, USA. Several of the scholars involved in the evolution of the V-Dem Dataset have contributed to the literature on democracy measurement long before the V-Dem came to life (for example Michael Coppedge, Carl Henrik Knutsen, Jan Teorell or Pamela Paxton to name a few). They were well aware of the ongoing debate about the definition, scaling and aggregation of existing democracy measures. As a result the V-Dem Dataset provides answers to several of the problems discussed in the first section and is probably the most stringent and transparent Dataset on democracy available today.

A first version of the dataset was introduced in Lindberg et al. (2014). The most recent version of data (the one used in this paper) is Version 7.1, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Bernhard, Fish, Glynn, Hicken et al. (2017) The data is available in country-year as well as in country-date format. Numerous country experts, coordinators and research assistants code basic variables capturing distinct features of a democratic system.<sup>68</sup> The variables are distinguished by type: type A variables are "based on extant sources and (...) factual in nature"<sup>69</sup> while type B and C variables are coded by country coordinators and experts. Type B variables are more factual items than C and do not contain as much judgment. Type C variables on the other hand require a greater amount of judgment and country specific knowledge, such as language and state of affairs. The question of

<sup>65</sup>As criticized by Munck & Verkuilen (2002) p. 25; Munck (2009), p. 50 - 51.

<sup>66</sup>See Cheibub et al. (2010), p. 75.

<sup>67</sup>See House (2017b).

<sup>68</sup>For more information on the variable types and coding procedure please refer to Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017) pp. 36.

<sup>69</sup>See Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 36.

how to aggregate variables coded by multiple experts into a single "best estimate" for each variable is quite important in this context (for example when some of the expert coded questions are coded on an ordinal scale). Most of the (C) variables are based on questions with answers on an ordinal scale and are thus aggregated across coders using Bayesian item response theory models.<sup>70</sup> Once the basic variables are aggregated into indices they are assigned type D.<sup>71</sup>

One of the main qualities distinguishing the V-Dem Dataset from others is their definition and conceptualization of democracy: V-Dem acknowledges the fact that a democracy measure's validity/its concept-measure consistency hinges on the proper definition of the underlying concept. Given the multifaceted nature of the concept of democracy, they provide disaggregate data. This gives empirical scholars the opportunity to construct democracy measures based upon concepts defined as needed in particular research frameworks. As a potential starting point, they propose considering democracy as a multidimensional concept consisting of the following five distinct dimensions:

1. The electoral dimension<sup>72</sup>
2. The participatory dimension<sup>73</sup>
3. The egalitarian dimension<sup>74</sup>
4. The deliberative dimension<sup>75</sup>
5. The liberal dimension<sup>76</sup>

In addition, V-Dem also recognizes the importance of an aggregation procedure reflecting the theoretical relationships between the concept's attributes for concept-measure consistency. Low level indices (D-type) in the V-Dem Dataset are combined into several mid-level indices, which in turn are then aggregated into high level indices reflecting these five dimensions of democracy. In this aggregation process issues of complementarity/substitutability or family resemblance between these dimensions are addressed. In the V-Dem definition of democracy, the electoral dimension is circled out as the core element without which no country shall be labelled democratic.<sup>77</sup> Hence, the Electoral Democracy Index is combined with the high level indices of dimensions 2 - 5 to create four indices of democracy.<sup>78</sup> These five indices are considered to embody the "varieties of democracy" (V-Dem).

Detailed information regarding V-Dem disaggregate data is given in the V-Dem Codebook, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017). The basic A,B,C-type variables are aggregated into low level indices (D-type), for example by "adding a denominator (e.g., per capita), by creating a cumulative scale (total number of...), or by aggregating larger concepts (e.g., components or indices of democracy)", Coppedge, Gerring, Lindberg, Skaaning,

---

<sup>70</sup>The measurement models are described in detail in Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), pp. 29. Marquardt & Pemstein (2018), further discuss item response theory models and compare their performance.

<sup>71</sup>More detailed information on the variable types can be found in Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), p. 17 - 18.

<sup>72</sup>"(...) embodies the core value of making rulers responsive to citizens through competition for the approval of a broad electorate during periodic elections", Lindberg et al. (2014), p. 160.

<sup>73</sup>"(...) embodies the values of direct rule and active participation by citizens in all political processes; it emphasizes nonelectoral forms of political participation such as through civil society organizations and mechanisms of direct democracy.", Lindberg et al. (2014), p. 160.

<sup>74</sup>"(...) holds that material and immaterial inequalities inhibit the actual exercise of formal rights and liberties; hence a more equal distribution of resources, education, and health across socioeconomic groups should enhance political equality.", Lindberg et al. (2014), p. 161.

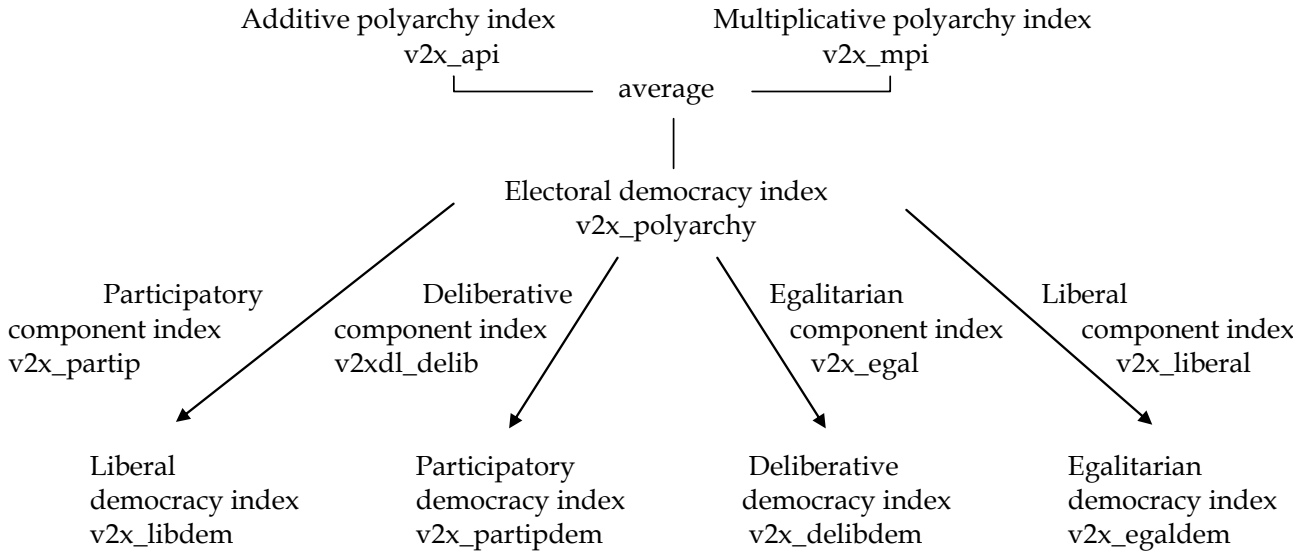
<sup>75</sup>"(...) enshrines the core value that political decisions in pursuit of the public good should be informed by respectful and reasonable dialogue at all levels rather than by emotional appeals, solidary attachments, parochial interests, or coercion.", Lindberg et al. (2014), p. 160.

<sup>76</sup>"(...) embodies the intrinsic value of protecting individual and minority rights against a potential "tyranny of the majority." This is achieved through constitutionally protected civil liberties, strong rule of law, and effective checks and balances that limit the use of executive power.", Lindberg et al. (2014), p. 160.

<sup>77</sup>See Lindberg et al. (2014), p. 161.

<sup>78</sup>The Liberal Democracy Index, v2x.libdem, the Participatory Democracy Index, v2x.partipdem, the Deliberative Democracy Index, v2x.delibdem as well as the Egalitarian Democracy Index, v2x.egaldem.

Figure 1: Aggregation process for high level V-Dem Indices



Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 36.

The Electoral Democracy Index, Polyarchy, reflects the minimalist democracy definition of contestation and participation.<sup>79</sup> To construct it, five low level indices reflecting freedom of expression, association, suffrage and free and fair elections are aggregated into two mid level indices of electoral democracy: the additive and the multiplicative polyarchy index. The additive index is constructed by weighted addition of the five low level indices and reflects the substitutability of democracy’s attributes. The multiplicative polyarchy index on the other hand is constructed by multiplying the five low level indices and captures the complementarity of the five concepts (a very low score on one of the components will lower its’ overall multiplicative polyarchy index). The Electoral Democracy Index, joins complementarity and substitutability by averaging the additive and multiplicative polyarchy indices. The aggregation process for the other four high level democracy indices is analogous. It is displayed in Figure 5.

The arrows represent an aggregation procedure which is averaging multiplication (complementarity) and addition (substitutability) of the two components:

$$\text{democracy index} = \frac{1}{4} \cdot (\text{polyarchy}^{1.6} + \text{component index}) + \frac{1}{2} \cdot \text{polyarchy}^{1.6} \cdot \text{component index}$$

Polyarchy influences each high level democracy index to the power of 1.6.<sup>80</sup> Setting a higher rate of influence for Polyarchy than any of the other component indices underlines the importance of the electoral democracy principle. Contestation and participation should be satisfied to a certain degree before further aspects of democracy can be employed to distinguish between higher levels of democracy.

The V-Dem Indices are continuous and range between [0,1]. Since in some instances ordinal measures might be needed, the V-Dem Project also offers the main indices as ordinal variables with 3, 4 or 5 categories respectively. The classification rules for the ordinal indices are also provided, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), pp. 268.

<sup>79</sup>The Online Appendix contains a figure displaying the components and aggregation rule for Polyarchy.

<sup>80</sup>The derivation of this rate is explained in the Online Appendix as well as in Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), p.10.

In addition to the democracy indices the V-Dem Dataset provides measures of uncertainty for each of the indices. For each of the indices the standard deviation (index suffixed `_sd`) as well as the "interval in which the measurement model places 68 percent of the probability mass for each country-year score"<sup>81</sup> (denoted by index suffixed `_codelow` and `_codehigh`) are given. These highest posterior density intervals can be seen as an indication of the skewedness of the underlying distribution: if the distances between the point estimate and the upper and lower bound are not equal, the underlying posterior distribution is skewed.

The dataset is very transparent. Even the project manager, responsible for crafting a specific variable, or the compiler is listed in the data. The number of experts coding a variable is also given (variables suffixed with `_nr`). The default number of coders for the period 1900-2012 is 5 or more.

The project documentation is extensive. Aside from the Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), the project for example offers documentation on methodology, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), on the units of observation, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Ciobanu & Saxer (2017), as well as on the project's setup, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Mechkova, Pernes, Olin, Saxer, Stepanova & Roemer (2017).

To ensure consistency of codings accross countries and over time bridge and lateral codings are employed.<sup>82</sup> Bridge coding refers to one coder coding a certain questions for multiple countries over the entire time series while lateral coding refers to the coding of a variable across all countries in a single year. Acknowledging the fundamental role of the country experts V-Dem provides information on country expert selection processes.<sup>83</sup> To avoid repressions the contry experts remain anonymus.<sup>84</sup>

An important feature of the V-Dem data to note is its sensitivity due to the Bayesian item response theory models. With addition of "new information", for example through the addition of a new bridge or lateral coder or the addition of a new year the point estimates for the indices change from each dataset version to the next. As a result country scores differ between the versions. While this might seem odd at first, this sensitivity to new information is one of the biggest strengths of the data: While Polity2 is struggling with the lack of variation in high democracies and low autocracies (as mentioned in section Polity Index) the above mentioned sensitivity of the V-Dem data will facilitate its adaption to new international constellations without loss of variation. Furthermore, it highlights that the country scores are point estimates with inherent uncertainty. This is a crucial point to keep in mind especially when creating rankings based on democracy measures.<sup>85</sup>

One minor issue to consider with the V-Dem dataset is that their distinction between de facto and de jure aspects of a polity (and hence their influence in the democracy measures) is not discussed or illustrated. Polyarchy, for example, contains a suffrage requirement and thus a measure of institutionalized enfranchisement (de jure participation). However, it is questionable how far actual (de facto) participation is captured. The data introduced in Vanhanen (2000) as well the Scalar Index of Polities by Gates et al. (2006), for example, measure participation as the percentage of the population which actually voted in the most recent parliamentary or presidential election (or both).<sup>86</sup> Considering voter turnout as a sole measure of participation can produce misleading results, for example if voting is mandatory. To capture a de facto aspect of participation Polyarchy includes an indicator

---

<sup>81</sup>Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 37.

<sup>82</sup>See Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), pp. 25.

<sup>83</sup>See section on country expert recruitment in Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), p.18.

<sup>84</sup>See Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), p.21.

<sup>85</sup>See Høyland et al. (2012) for a thorough discussion of the pitfalls of disregarding uncertainty.

<sup>86</sup>See Vanhanen (2000), p. 253, and Gates et al. (2006), p. 897.



Table 6: Pairwise correlation coefficients, bottom left: for all observations, top right: only observations coded in all datasets. Number of observations in parenthesis below.

	Polity2	Polyarchy	FHI
Polity2	1	0.9083 (6,546)	0.8889 (6,546)
Polyarchy	0.8661 (11,781)	1	0.9219 (6,546)
FHI	0.8892 (6,580)	0.9200 (6,902)	1

on electoral violence.<sup>87</sup> A discussion of the inclusion of de jure/de facto criteria in the definition of democracy would be desirable.

Nevertheless, with the public availability of disaggregate and aggregate data, theoretical justification for the detailed aggregation rule and comprehensive spatial and temporal coverage, the V-Dem dataset provides the most well-documented and well-grounded collection of democracy measures available today.

### 3 Comparison

The previous section outlined the diverse definitions and aggregation procedures embedded in each of the democracy measures. In addition, the three datasets vary considerably in terms of countries and years covered. These differences in definition, in availability of disaggregate data and country-years coded by each index are substantial and often prove decisive with respect to the question which index to use in which framework. To assess dissimilarities in coding between the three indices the sample is restricted to those observations available in all datasets for the remainder of this section. This sample will be referred to as the trunk dataset.

Note that while the reduction to a trunk dataset is necessary it is also a very harsh restriction and a considerable amount of information is disregarded. For example, restricting the sample to countries included in the Freedom House data limits the timeline to post-1972 years. However, Polity2 and V-Dem Data can be compared using data from 1900 on. It becomes even more drastic when the observations left out are chosen by attribute (and not by year): In the Polity2 coding scheme observations during which a country is occupied by foreign powers constitute missing values (which is not the case for Polyarchy and FHI). Thus, these cases are not part of the trunk dataset.

How does this restriction affect similarities between the measures? Comparing pairwise correlations<sup>88</sup> to the correlations obtained using the trunk dataset (see Table 6) shows that for the observations in the trunk dataset (top right) the indices' correlation is larger or (almost) equal to the pairwise comparison (displayed in the bottom left part).

Coppedge et al. (2011)<sup>89</sup> point out that high intercorrelations between the democracy indices are at least partly due to observations which are "perfectly" democratic/autocratic. The trunk dataset contains 6,546 observations for 167 countries from 1972 to 2015. In comparison to the period 1900 to 1971 the number of "perfect" democracies drastically increased after 1972, which could explain the higher intercorrelations in the trunk data. As displayed in Table 7 around 21.2% of the observations in the

<sup>87</sup>Election other electoral violence, v2elpeace, see Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 97.

<sup>88</sup>Note, that the V-Dem Data starts in 1900. When computing pairwise correlation coefficients thus the number of observations is much higher than when doing so for the trunk dataset.

<sup>89</sup>See Coppedge et al. (2011), p. 252.

Table 7: Number of perfect democracies/autocracies (+/-10 in the Polity2 coding scheme) in the period 1900 - 1971 and the trunk dataset (1972 - 2015)

Number of perfect...	1900-1971	1972 - 2015
... autocracies	265	194
... democracies	1,026	1,194
Total	1,291	1,388

Table 8: Pairwise correlation coefficients for observations which are not coded "perfect" democracies/autocracies (10/-10) in Polity2.

Bottom left: for all observations, top right: only observations coded in all datasets. Number of observations in parenthesis

	Polity2	Polyarchy	FHI
Polity2	1	0.8648 (5,158)	0.8423 (5,158)
Polyarchy	0.8002 (9,169)	1	0.8710 (5,158)
FHI	0.8425 (5,173)	0.8710 (5,158)	1

trunk dataset are coded as perfect democracies/autocracies (+/- 10 on the Polity Scale). When removing them from the sample the correlations decrease as displayed in Table 8. Limiting the dataset to observations available for all three indices thus implies obtaining a sample with a high fraction of perfect democracies/autocracies, which in turn contributes to an intercorrelation between indices which is larger or (almost) equal to the pairwise comparisons.

### 3.1 Summary statistics

Table 9 displays summary statistics for the democracy indices discussed above. Since it represents the minimalist democracy definition of contestation and participation the Polyarchy Index<sup>90</sup> was chosen from the V-Dem indices.<sup>91</sup> For facilitated comparability the FHI was reversed and both, FHI and Polity2 were normalized between 0 and 1.<sup>92</sup> Due to different geographical and temporal scopes covered by each index the number of observations for which the summary statistics are computed are quite different.<sup>93</sup> The summary statistics for all observations available per index are displayed on the top part of Table 9. The summary statistics for the trunk dataset are displayed in the lower part of Table 9.<sup>94</sup>

<sup>90</sup>Polyarchy was used exactly how it is provided in the data, i.e. it was not rescaled or normalized.

<sup>91</sup>Summary statistics for the other V-Dem democracy indices are given in the Online Appendix.

<sup>92</sup>

$$\text{Normalized Polity2} = \frac{\text{Polity2} + 11}{21}, \text{ reversed and normalized FHI} = \frac{7 - \text{FHI}}{6}$$

<sup>93</sup>More information on the different geographical and temporal scopes covered by Polity2 and Polyarchy can be found in Boese & Kamin (2018).

<sup>94</sup>The Online Appendix includes a list of countries and their respective years coded in the trunk dataset.

Table 9: Summary statistics for the democracy indices

Summary statistics for...						
...all observations available per index						
Variable	Obs.	Mean	Median	St. D.	Min	Max
Polity2	16,992	0.4724	0.35	0.3536	0	1
Polyarchy	17,036	0.3179	0.2055	0.2788	0.0086	0.9471
FHI	6,936	0.5071	0.5	0.3375	0	1

...the trunk dataset						
Variable	Obs.	Mean	Median	St. D.	Min	Max
Polity2	6,546	0.5689	0.7	0.3647	0	1
Polyarchy	6,546	0.4545	0.4061	0.2868	0.0140	0.9471
FHI	6,546	0.5053	0.5	0.3362	0	1

The distributions of all three measures are u-shaped,<sup>95</sup> with peaks (particularly pronounced for Polity2 and FHI) at the extremes.<sup>96</sup> As discussed above 18.2% of the observations in the trunk dataset are coded as "perfect" democracies by Polity2. This high number of perfect democracies is also reflected by a Polity2 median which is substantially larger than Polity2's mean. The distributions of FHI and Polyarchy are more even. However, FHI considers around 14% of observations in the trunk dataset as perfect democracies and around 9% as perfect autocracies while no observation is coded as either by the V-Dem Project. Polity2 and FHI capture very little variation in highly democratic/autocratic systems (the US, for example, is coded as a perfect democracy from 1972 to 2015) as opposed to Polyarchy.<sup>97</sup> This has an important implication for the choice of measure in frameworks in which highly democratic/autocratic countries are examined. The FHI also has a comparatively high number (around 8.5% of the observations) of perfect autocracies, i.e. observations coded as 0. In addition, the high and low numbers of perfect democracies/autocracies have a noteworthy theoretical implication for the future development of the respective countries' time series: for those countries the system of governance is not able to improve/deteriorate.

## 3.2 Examining the differences

The following section examines the differences in coding between the three democracy measures and their implication for the country ranking within each index. The section closes with a country study exemplifying the differences in coding variation, in ranking as well as in disaggregation possibilities.

### 3.2.1 Difference with respect to attributes of the democracy definition

Polity2 assigns a value of 0<sup>98</sup> for observations in which a country undergoes a period of interregnum/anarchy and prorates the respective country's democracy value over the length of a transition period (see section *Polity Index*).

The boxplot in Figure 2 shows that Polyarchy and FHI are both much lower than Polity2 (i.e. the interquartile range of the differences is strictly positive) for countries in anarchy and transition periods.

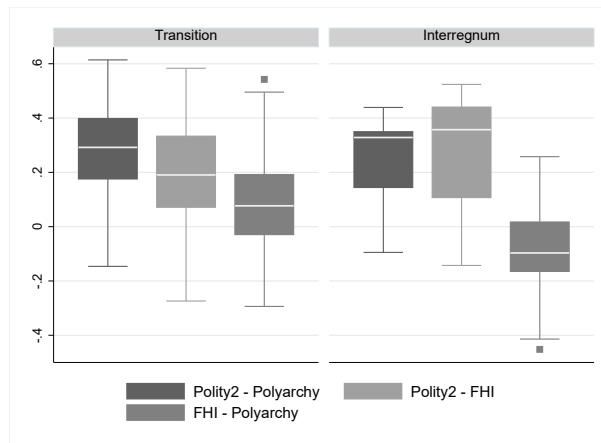
<sup>95</sup>Several studies, such as Goertz (2006), discuss how colinearity of the indicators forming the democracy indices can contribute to this u-shape.

<sup>96</sup>The Online Appendix provides a histogram as well as a table with percentiles for the three democracy measures in the trunk data giving further insight into their empirical distribution.

<sup>97</sup>The Online Appendix includes an exemplary discussion of the within and between country variation in the Polity2 data.

<sup>98</sup>A Polity2 value of 0 translates into  $11/21 \approx 0.52$  on the normalized polity scale (between 0 and 1).

Figure 2: Boxplot of differences between index values for cases coded as transitions or interregnum in the Polity Scheme



Interestingly, at the same time the difference in values between FHI and Polyarchy is low for these cases (median differences of around  $\pm 0.1$ ), while the differences between each of them and the Polity2 coding are substantial. Countries undergoing periods of anarchy or transition appear to be systematically overrated by Polity2. In research frameworks in which such countries play an important role FHI or V-Dem indices should therefore be preferred over Polity2.

As mentioned in section Polity Index the original aim of the Polity data was to capture a country's formal institutional setup. This led to the Polity Index reaching the boundaries of its domain in cases where there were no regular institutions, i.e. in cases of interregnum, transition or occupation. Even though V-Dem includes several aspects beyond the formal institutions the absence of a regular institutional framework in post WWII-Germany seems to be posing difficulties for V-Dem as well: Germany is not coded between 1946 - 48. Since this is the only such case the recommendation to prefer V-Dem indices or FHI whenever a share of the relevant observations falls in categories discussed above remains valid.

### 3.2.2 Difference in coding by level of democracy

The high intercorrelations between the democracy measures discussed above show that there is a strong relationship between them. This is to be expected, as the three indices are supposed to measure the same thing - democracy. But, the correlations do not provide information on the "agreement" between the indices. Agreement can be thought of as the identity line, when two indices are plotted with respect to each other. If all observations are close to the identity line, the indices "agree" and the differences between them are close to zero. A case with high correlation, but limited agreement occurs, for example, if one index were to consistently code each country half as democratic as the other.

In the following, the general agreement between the indices is examined and systematic differences are explored. Let's assume there was a "hypothetical democracy scale" along which all countries could be sorted. Does one index code highly democratic/autocratic countries (on the top/bottom of the hypothetical scale) systematically different from another index? The "hypothetical democracy scale" is, of course, unknown (and given the different definitions underlying the democracy measures it is strictly hypothetical). However, assuming the three democracy measures are three ways of capturing very similar concepts, we can use the average of the three measures as a proxy for the hypothetical scale. Differences between the index pairs are plotted against this average, i.e. the "hypothetical democracy scale" in Figure 3. The solid line at  $y=0$  marks the observations for which there is perfect agreement between the indices (i.e. where the difference between them equals zero). The dashed lines mark the 95th-percentile and the 5th-percentile, hence, 90% of the observations fall between the

Figure 3: Pairwise differences between indices plotted over average democracy levels, horizontal lines at 0, the 95th-percentile and the 5th-percentile

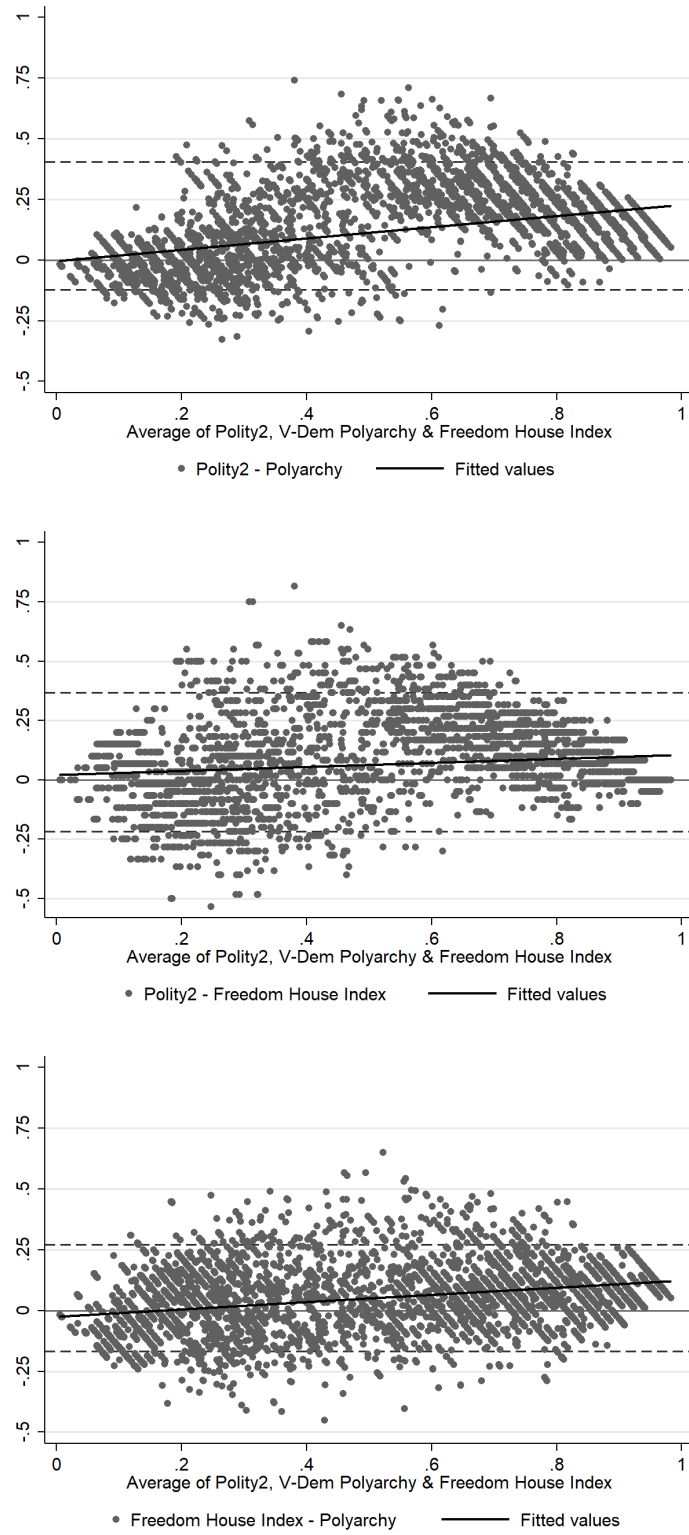


Table 10: Results of ttest of mean differences between indices

Difference between...	Mean
Polity2 - Polyarchy	0.1144***
Polity2 - FHI	0.0636***
FHI - Polyarchy	0.0508***

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.10$

top and bottom dashed line. The black line represents a linear fit (pooled OLS) of the differences in democracy measures on "hypothetical democracy".<sup>99</sup>

For all indices the differences get closer to zero towards the democratic/autocratic extreme. The plot for Polity2 and V-Dem Polyarchy Index provides useful insight on the agreement between the two indices: The fitted line is positive and increasing. In a large number of cases Polity2 rates countries more democratic than the Polyarchy Index. For the intermediate 90% of observations the difference between the two index values tends to be positive. In fact, for countries in the upper half of the democracy scale, there is very little agreement: The difference in index values becomes larger as the level of "hypothetical democracy" increases. The vast majority of countries located between 0.55 and 0.8 on the hypothetical democracy scale is coded strictly more democratic by Polity2 than by Polyarchy Index. For countries at the very low end (until 0.1) of the democracy scale, on the other hand, the two indices seem to agree to some extent: differences are symmetrically distributed close to 0. There are comparatively few observations for which Polyarchy is significantly larger than Polity2 (those in the 5-th percentile). They occur for countries on the lower half (between 0.1 and 0.7) of the hypothetical democracy scale.

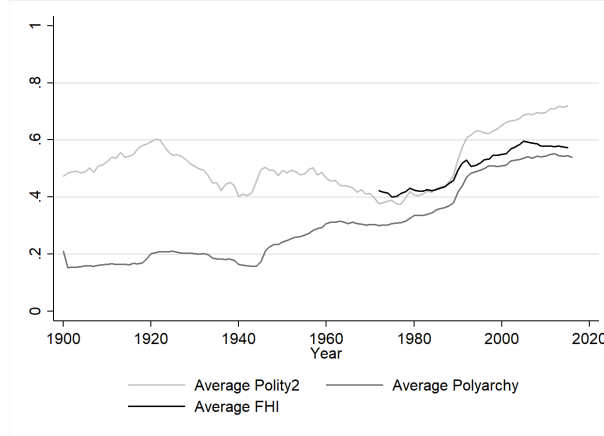
The plot for FHI and Polyarchy also displays an increase in difference between the two measures with rising hypothetical democracy values. The fitted line is increasing, but it is negative for hypothetical democracy values below 0.16. Countries on the lower end of the hypothetical democracy scale are coded slightly more democratic by Polyarchy than by FHI. However, this is reversed for countries above the 0.18 threshold: for those FHI tends to be slightly larger than Polyarchy and increasingly so with rising levels of democracy. Large differences (as measured by the 10% of observations for which the absolute differences in democracy values are particularly high) occur for countries anywhere on the democratic scale except the end points.

Polity2 and FHI's agreement is almost consistent across hypothetical democracy values: The fitted line is nearly horizontal, but positive. Polity2 tends to code countries slightly more democratic anywhere on the democracy scale. Examining the large differences only, FHI tends to be larger than Polity2 only for countries in the lower half of the democratic scale.

The assessment of dis-/agreements between democracy measures concludes with testing whether there are significant differences in coding between the indices: is the mean difference between the indices is different from zero ( $H_0$ )? The results are displayed in Table 10. They are in line with the findings from Figure 3 above. The mean differences are positive and significantly different from zero. On average Polity2 assigns the highest democracy values, followed by FHI. Polyarchy on average assigns the lowest democracy values. Due to Polyarchy's definition this comes as no surprise: as discussed above it captures the minimalist democracy definition of contestation and participation.

<sup>99</sup>Detailed regression results are provided in the Online Appendix.

Figure 4: World averages of Polyarchy, Polity2 and FHI



### 3.2.3 Difference by region

To examine the results from the last section world averages from 1900 - 2016 are plotted in Figure 4. Recall, that the Polyarchy Index captures the minimalist democracy definition of contestation and participation. When examining the world averages this becomes obvious as it is - on average - quite austere, i.e. lower than FHI or Polity2, in its coding. Interestingly, the world averages show that FHI values fluctuate between Polity2 and Polyarchy only from 1990 on-wards. Before that the FHI average is above the Polity2 one.

Figure 5 breaks up the dataset by geopolitical region<sup>100</sup> and displays regional averages for each of the democracy measures. For most regions the Polity2 values are larger or equal (almost equidistant) to the Polyarchy values while the FHI mostly "fluctuates" between the other two. The change in coding between FHI and Polity2 averages noted in Figure 4 is particularly visible in the following regions: In Latin America & Caribbean, Middle East & North Africa as well as East and South Asia the Polity2 average is below the FHI average prior to 1990, but above it after. In addition, when examining the regional averages for Eastern Europe & Central Asia Polity2 and Polyarchy Index show a high level of agreement for the observations before 1990. The Freedom House Index on the other hand codes the Eastern European observations much less democratic than the other two indices between 1972 - 1990. This might be a reflection of the Freedom House Index' early ideological bias discussed in section *Freedom House*. It could also indicate Polyarchy values which are too high. It is the only time span and region in which Polyarchy is higher than both, Polity2 and FHI. In the following section, the coding of the Eastern European & Central Asian countries will be examined in more detail to shed light on this finding.

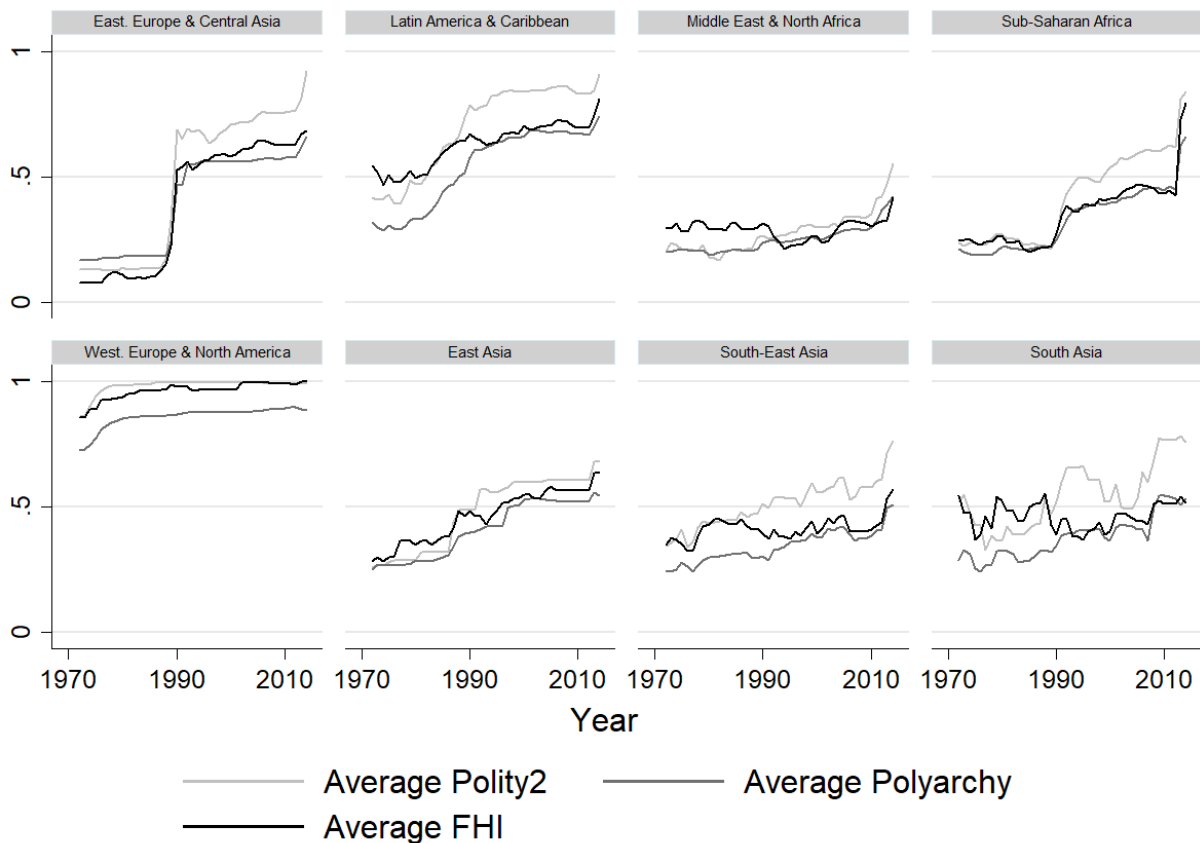
### 3.2.4 Eastern Europe & Central Asia

The regional comparison in the Figure 5 provides some insight into the "austerity" of each democracy index. On average Polity2 assigns the highest and Polyarchy the lowest values while the FHI fluctuates between the other two. The only time and region in which there seems to be a systematic deviation from this scheme is in Eastern Europe & Central Asia (EECA) before 1990. There are 31 countries in the region.<sup>101</sup> Albania, Bulgaria, Mongolia and Romania provide good examples of the coding phenomenon discussed above. Hence, their coding will be studied in more detail in the remainder of this section. The respective democracy values assigned by Polity2, FHI and Polyarchy

<sup>100</sup>A detailed list of countries and their respective regions can be found in the Online Appendix.

<sup>101</sup>A list of these countries is provided in the Online Appendix.

Figure 5: Regional averages of the three democracy measures



are displayed in Figure 6.<sup>102</sup> To understand why Polyarchy codes each country comparatively high Polyarchy is broken up into its components in Figure 7. The share of population with suffrage as well as the Elected officials index<sup>103</sup> are coded with the highest possible value over the entire time span. This strongly contributes to the high Polyarchy values in these countries prior to 1990. It is a central difference between the three indices and their underlying definitions of democracy. As discussed above, Polity2 and FHI<sup>104</sup> do not include any components capturing suffrage requirements. Polity2's categories XROPEN ("Recruitment of the chief executive is "open" to the extent that all the politically active population has an opportunity, in principle, to attain the position through a regularized process", Marshall et al. (2017a), p. 22.) and XRCOMP ("Competitiveness refers to the extent that prevailing modes of advancement give subordinates equal opportunities to become superordinates", Marshall et al. (2017a), p. 21.) are closest in meaning to V-Dem's elected officials index. Here a lack of detail and organization in theoretical concept of the Polity2 scheme becomes apparent: the very definition of XROPEN mixes several important attributes, namely the regularized electoral process (in V-Dem captured by the elected officials index), the extent to which the politically active population is *de jure* allowed to participate (i.e. the share of population with suffrage) as well as the actual *de facto* opportunities the population has to influence the electoral process (in V-Dem this is covered, for example, through the clean elections index). The Polity2 categories are not distinguishing between

<sup>102</sup>The Online Appendix contains a Figure displaying the codings for all other countries in the EECA region. See Figure 16.

<sup>103</sup>The elected officials index, v2x\_elecoff addresses the question "Is the chief executive and legislature appointed through popular elections?", Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 59.

<sup>104</sup>Since no disaggregate data is available for FHI and the checklist questions are vague at best, its coding of the EECA region cannot be discussed further here.



Figure 6: Democracy coding by Polity2, FHI and Polyarchy for Albania, Bulgaria, Mongolia and Romania

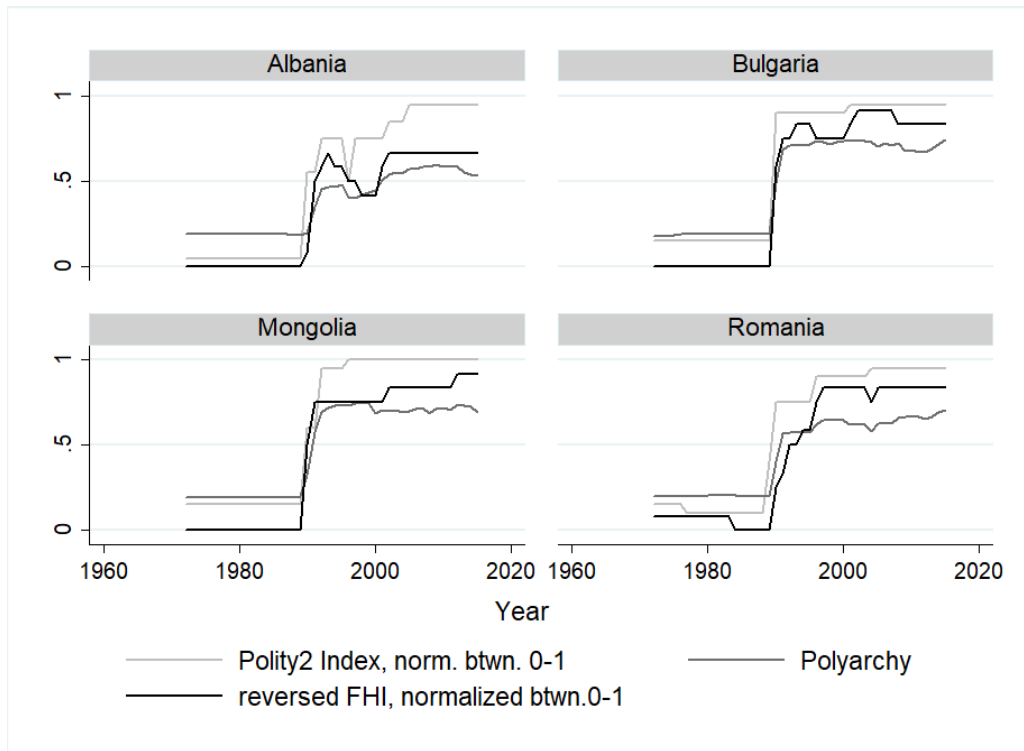


Figure 7: Components of Polyarchy for Albania, Bulgaria, Mongolia and Romania

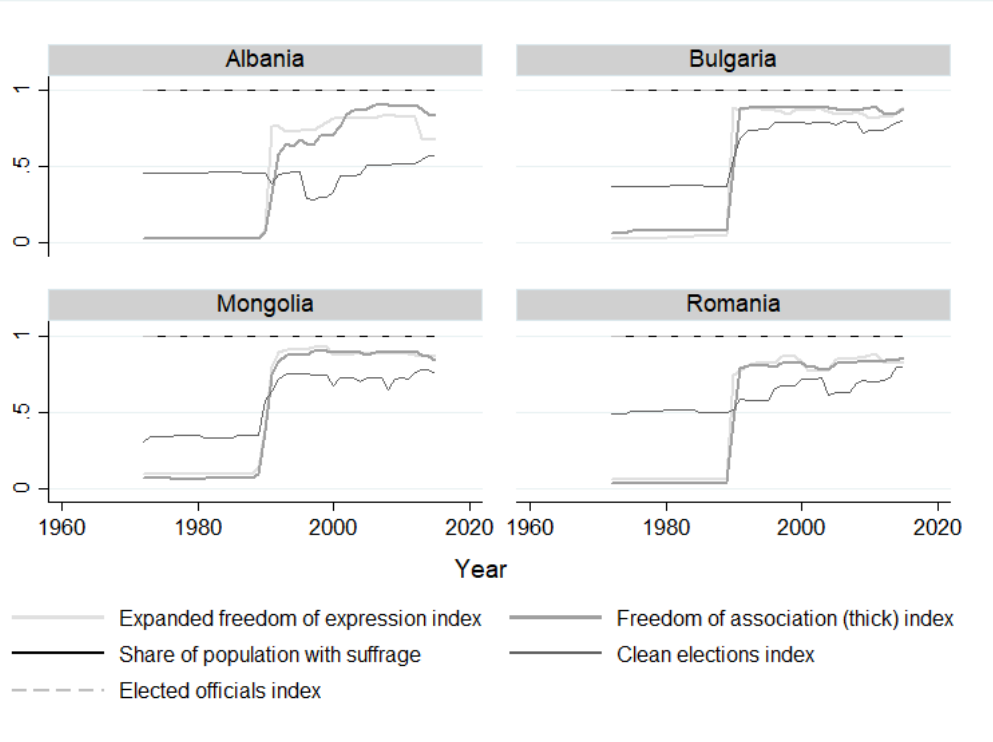
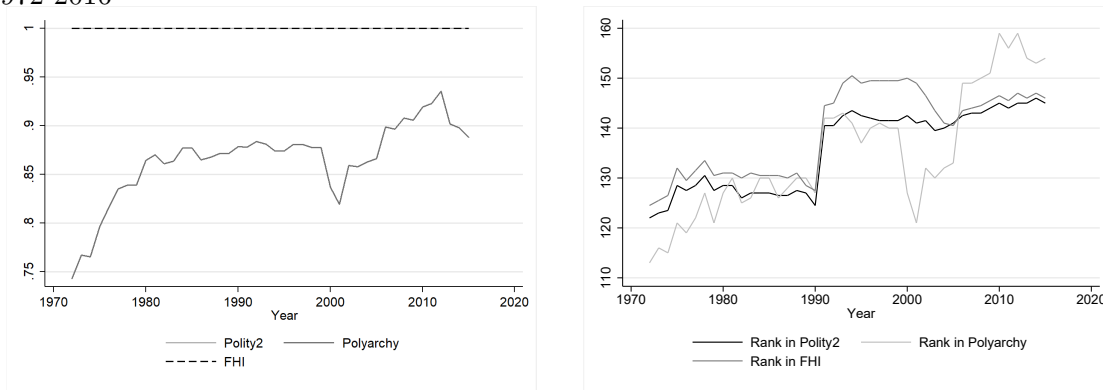


Figure 8: Democracy index values (left) and yearly rankings (right) for the United States of America from 1972-2016



different attributes of the underlying concepts of the de jure electoral process and the de facto as well as de jure participation of the population.

The Online Appendix contains a section comparing the country ranking of each of the three democracy indices. It shows that the countries are ranked almost consistently across all regions except for EECA prior to 1990. The difference in the democracy definition covered (by the inclusion of suffrage) is a major driver behind the comparatively high values of Polyarchy in that time and region. It ultimately also leads to a different ranking for the corresponding countries. In conclusion, a closer examination of the EECA region demonstrates that differences in the underlying democracy definition can and do lead to differing country ratings. In applications covering time periods and regions such as EECA prior to 1990 this can potentially be a source of results varying with the democracy measure used.

### 3.2.5 United States of America

A closer examination of the timeseries of United States will illustrate differences in temporal variation, in ranking and the disaggregation possibilities between the three indices.

Figure 8 displays the democracy index values and yearly rankings of the United States between 1972 and 2016. The US is coded as a perfect democracy receiving the highest possible value throughout the entire time series by both, Polity2 and FHI. This lack of variation for the two indices is also found in most highly democratic/autocratic countries.<sup>105</sup> In circumstances in which the research question at hand concerns such countries this feature needs to be kept in mind.

The differences in ranking between FHI and Polity2 are minor. However, the difference to Polyarchy is rather large. From 1999 to 2001 US Polyarchy drops by about 5.8%. What caused this sharp decrease? Plotting the components of Polyarchy for the United States over time (see Figure 9) shows that the clean elections index,  $v2xel\_refair$ ,<sup>106</sup> is the main driving factor behind it.

<sup>105</sup>A table displaying the within and between country variation for Polity2 can be found in the Online Appendix.

<sup>106</sup>“Question: To what extent are elections free and fair? Clarifications: Free and fair connotes an absence of registration fraud, systematic irregularities, government intimidation of the opposition, vote buying, and election violence.”, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 58.

Figure 9: Components of Polyarchy for the United States over time

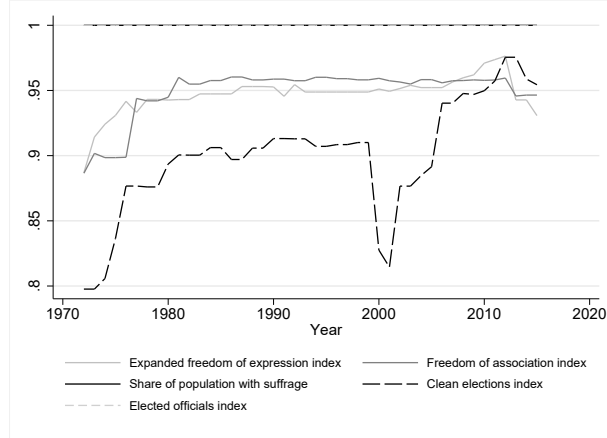
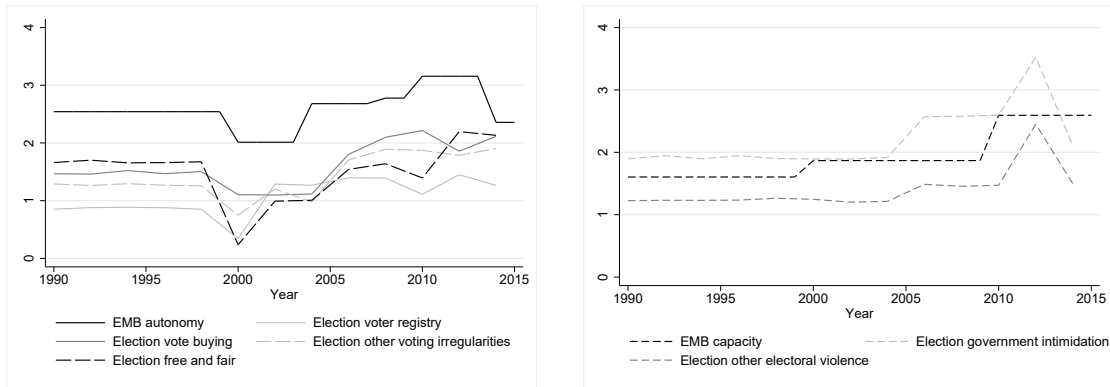


Figure 10: Subcomponents of the clean elections index for the United States of America from 1990 - 2016



Further decomposing the clean elections index into its subcomponents<sup>107</sup> (see Figure 10) exposes EMB autonomy (v2elembaut),<sup>108</sup> election voter registry (v2elrgstry),<sup>109</sup> election vote buying (v2elvotbuy),<sup>110</sup> election other voting irregularities (v2elirreg),<sup>111</sup> and election free and fair (v2elfrfair)<sup>112</sup> as driving factors (see left graph).

In sum, the drop in Polyarchy (and subsequently in the ranking of the US in comparison to other countries in the year 2000) is due to the 2000 presidential election and is not captured by Polity2 or FHI. This difference in the US- values and ranking for each of the indices illustrates how Polyarchy is more sensitive than Polity2 or FHI, which is both, an advantage (higher variability, more information is included) as well as (slight) flaw of the index. As discussed in section *Varieties of Democracy (V-Dem) Dataset* V-Dem does not provide a thorough discussion of the de facto and de jure elements

<sup>107</sup>Note that most of these indicators only occur in election years and are then repeated over election regime periods.

<sup>108</sup>”Question: Does the Election Management Body (EMB) have autonomy from government to apply election laws and administrative rules impartially in national elections?”, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 86.

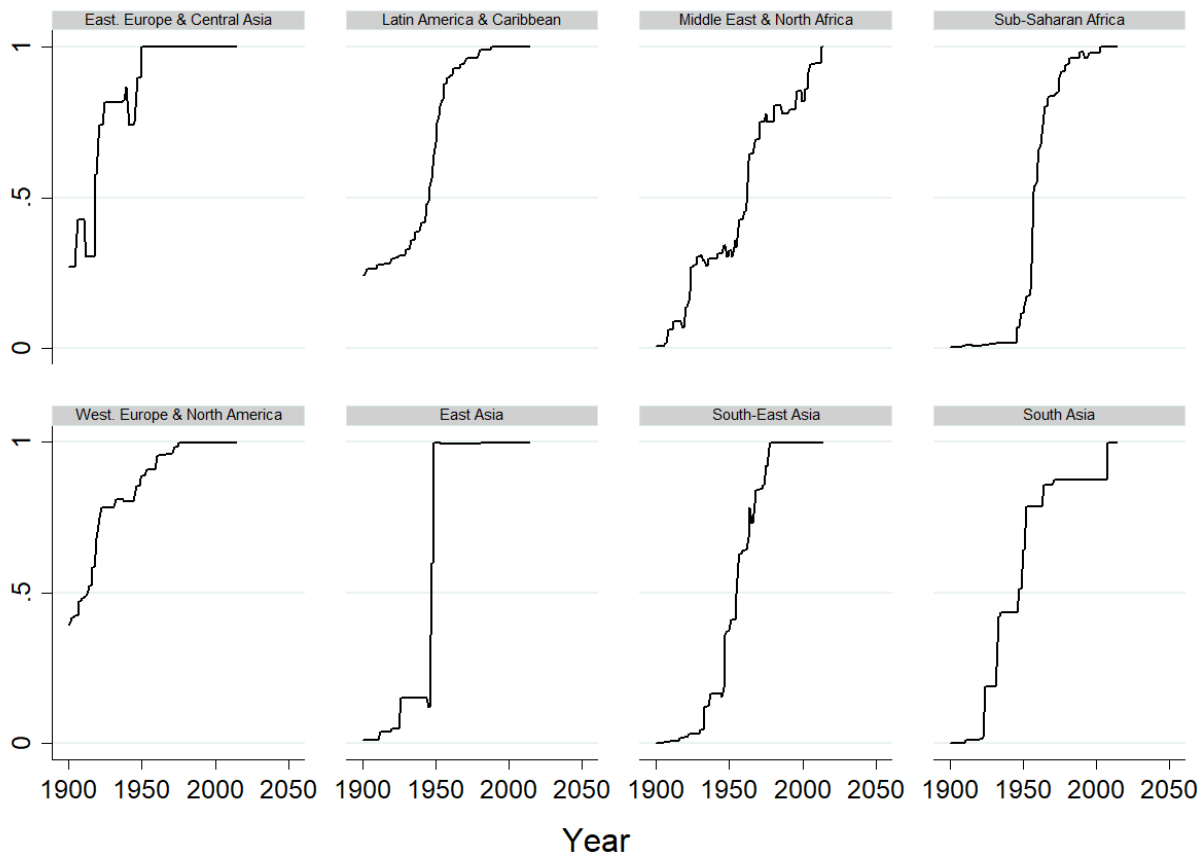
<sup>109</sup>”Question: In this national election, was there a reasonably accurate voter registry in place and was it used?”, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 90.

<sup>110</sup>”Question: In this national election, was there evidence of vote and/or turnout buying?”, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 94.

<sup>111</sup>”Question: In this national election, was there evidence of other *intentional* irregularities by incumbent and/or opposition parties, and/or vote fraud?”, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 95.

<sup>112</sup>”Question: Taking all aspects of the pre-election period, election day, and the postelection process into account, would you consider this national election to be free and fair?”, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p. 103.

Figure 11: Average regional shares of population with suffrage, 1900 - 2016. Source: V-Dem indicator v2x\_suffr



included in the democracy measure. De facto (as opposed to de jure) indicators by nature rely more on judgement calls and inherently suffer from an increased bias. The indicators driving the decrease in US Polyarchy values address the de facto aspect of the United States democracy in 2000. US Polyarchy drops from around 0.88 in 1999 to 0.84 in 2000. Whether this sharp decrease is an appropriate reflection of changes in US levels of democracy around the 2000 election or whether it might be influenced/amplified by ideological judgments inherent in de facto indicators can be debated. It comes down to the question "what share of the underlying democracy definition consists of de facto vs. de jure attributes of a polity?". None of the three democracy measures properly address this question.

Nevertheless, Polyarchy still outperforms the other two measures with respect to all points discussed in the paper. Polyarchy's increased sensitivity/variation (especially for highly democratic/ autocratic) countries can be a large advantage in a time-series cross-country setting over the countries which are coded as 1/0 for decades by Polity2 or FH. The one thing the V-Dem project could improve is to find a system of "checks and balances" for de facto attributes so as to not let this sensitivity get out of hand.

## 4 Discussion

The three democracy indices studied in this paper were introduced at different times and for different purposes. In section Definition of democracy the distinction between *de facto* and *de jure* attributes included in a democracy measure was discussed. It ties into the very definition of the underlying concept of democracy/authority patterns. The Polity data was created to examine the durability of institutional frameworks. Its aim was to capture the *de jure* framework. FHI on the other hand was

originally designed as a measure of civil liberties and political rights. It therefore is more of a *de facto* measure. V-Dem's Polyarchy is situated somewhat in between the two hypothetical *de facto* and *de jure* endpoints. There is no "optimal" partition of *de facto* and *de jure* elements to incorporate in a democracy measure. The research question of the project at hand and the corresponding perception of democracy should determine the levels of *de facto* and *de jure* attributes. None of the three democracy indices are "better" or "worse" because of their choice of partition. While the choice in itself does not affect the indices overall validity it does affect the values coded. The country studies provided in the previous section show that the inclusion of more or less of such *de facto* and *de jure* elements leads to different country assessments by the three measures. The example of the coding of the United States highlights how the inclusion of *de facto* attributes can be a main driver behind different ratings across indices. Empirical researchers planning to work with democracy measures hence need to contemplate what degree of *de facto* and *de jure* attributes is appropriate for their research.

Throughout the paper the loss of variation in particular for countries at the very top or bottom of the scales has been discussed. The two country studies in the previous section highlighted this for Polity2 and FHI in the United States as well as for the suffrage dimension in Eastern Europe & Central Asia. The latter example raises the question whether suffrage today is even an important indicator to include in a measure of democracy as there is little/no geographical variation. Figure 11 displays average shares of population with suffrage by region. The graphs show little or no variation in the last decade of the time series but a lot of variation in the first half of the 20th century. Thus the question posed above can be amplified in scope by asking: how should democracy measures adapt to the changing importance of certain dimensions of democracy over time? The share of population with suffrage has been an important dimension of democracy. It helped distinguish between different kinds of regimes and to capture some extent of the people's options to participate in the political process. Today, any new polity coming into existence will find it difficult to allocate suffrage to white adult males only, for example. The share of the population with suffrage is not as helpful in distinguishing between different polities today as it was 50 years ago. It goes well beyond the scope of this paper to provide a solution to the changing importance of certain democracy dimensions over time. Nevertheless, it is within the scope of this paper to point out that V-Dem excels at providing very precise and clear cut distinctions between a large number of dimensions. By providing the disaggregate data it also gives the empirical researcher the opportunity to include or disregard dimensions as needed.

## 5 Conclusion

This paper compares the three most commonly used democracy datasets: the PolityIV, the Freedom House and the V-Dem dataset. In a first step, it analyzes their respective democracy measures' underlying definition, their measurement scale, their data collection as well as the theoretical justification of their aggregation procedure to assess their validity and reliability. In addition, important strengths and weaknesses of each measure are discussed. The democracy indices from the V-Dem dataset surpass the other two democracy measures in all areas mentioned above. In a second step, this article compares the indices' respective coding of those observations included in all three datasets. There is a relatively high level of agreement in the country coding as well as in the country ranking between the indices for these cases. In conclusion, the most substantial differences between the indices lie in their coverage, the availability of disaggregate data and the above mentioned key areas. These are the central aspects for scholars to consider when choosing a democracy measure for their research.

Last but not least, the massive efforts undertaken and the equally enormous contributions of each of the three data projects discussed shall be underlined. As stated in previous sections each data project originated with quite different objectives and at different times. The first V-Dem dataset was released about 40 years after the first Polity and Freedom House data came out. In addition, it had been prepared by scholars familiar with the Polity and Freedom House data who knew the caveats and pitfalls of these datasets by heart. The V-Dem project was able to build their data on the foundations of scientific discourses about empirical democracy measurement since the introduction of Polity and Freedom House data. By offering extensive documentation, disaggregate data and sensitivity to new information through the Bayesian item response models the V-Dem data is optimally equipped to become the new standard in democracy measurement and to adapt to future challenges.

## References

- Acemoglu, D., Naidu, S., Restrepo, P. & Robinson, J. A. (2014), Democracy does cause growth, Technical report, National Bureau of Economic Research.
- Boese, V. A. & Kamin, K. (2018), ‘’tis but thy name that is my enemy: on the construction of macro panel datasets in peace economics’, *Economics of Peace and Security Journal* (**forthcoming**).
- Bogaards, M. (2010), ‘Measures of democratization: From degree to type to war’, *Political Research Quarterly* **63**(2), 475–488.
- Boix, C., Miller, M. & Rosato, S. (2013), ‘A complete data set of political regimes, 1800–2007’, *Comparative Political Studies* **46**(12), 1523–1554.
- Bollen, K. (1993), ‘Liberal democracy: Validity and method factors in cross-national measures’, *American Journal of Political Science* **37**(4), 1207–1230.
- Bollen, K. A. (1980), ‘Issues in the comparative measurement of political democracy’, *American Sociological Review* **45**(3), 370–390.
- Bollen, K. A. & Jackman, R. W. (1985), ‘Political democracy and the size distribution of income’, *American Sociological Review* **50**(4), 438–457.
- Bollen, K. A. & Jackman, R. W. (1989), ‘Democracy, stability, and dichotomies’, *American Sociological Review* **54**(4), 612–621.
- Bollen, K. A. & Paxton, P. (2000), ‘Subjective measures of liberal democracy’, *Comparative Political Studies* **33**(1), 58–86.
- Bush, S. S. (2017), ‘The politics of rating freedom: Ideological affinity, private authority, and the freedom in the world ratings’, *Perspectives on Politics* **15**(3), 711–731.
- Casper, G. & Tufis, C. (2003), ‘Correlation versus interchangeability: The limited robustness of empirical findings on democracy using highly correlated data sets’, *Political Analysis* **11**(2), 196–203.
- Cheibub, J. A., Gandhi, J. & Vreeland, J. R. (2010), ‘Democracy and dictatorship revisited’, *Public choice* **143**(1-2), 67–101.
- Coppedge, M., Alvarez, A. & Maldonado, C. (2008), ‘Two persistent dimensions of democracy: Contestation and inclusiveness’, *The Journal of Politics* **70**(3), 632–647.
- Coppedge, M., Gerring, J., Altman, D., Bernhard, M., Fish, S., Hicken, A., Kroenig, M., Lindberg, S. I., McMann, K., Paxton, P. et al. (2011), ‘Conceptualizing and measuring democracy: A new approach’, *Perspectives on Politics* **9**(2), 247–267.
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E. & Teorell, J. (2017), ‘V-dem comparisons and contrasts with other measurement projects’, *V-Dem working paper*. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2951014#](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2951014#) (accessed on Dec. 21, 2017) **2017**(45).
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Altman, D., Andersson, F., Bernhard, M., Fish, M. S., Glynn, A. et al. (2017), ‘V-dem codebook v7’, *V-Dem Reference Material* (2017). Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2968274](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968274) (accessed on Dec. 21, 2017) .
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Altman, D., Bernhard, M., Fish, M. S., Glynn, A., Hicken, A. et al. (2017), ‘V-dem dataset v7’, Available at: <https://www.v-dem.net/en/data/data-version-7-1/> (accessed on Dec. 21, 2017) .

- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Ciobanu, V. I. & Saxer, L. (2017), 'V-dem country coding units v7', *V-Dem Reference Material (2017)*. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2968285](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968285) (accessed on Dec. 21, 2017).
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Krusell, J., Marquardt, K. L., Mechkova, V., Pemstein, D., Pernes, J. et al. (2017), 'V-dem methodology v7', *V-Dem Reference Material (2017)*. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2968284](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968284) (accessed on Dec. 21, 2017).
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S.-E., Teorell, J., Krusell, J., Mechkova, V., Pernes, J., Olin, M., Saxer, L., Stepanova, N. & Roemer, J. v. (2017), 'V-dem organization and management v7', *V-Dem Reference Material (2017)*. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2968280](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968280) (accessed on Dec. 21, 2017).
- Crick, B. (2002), *Democracy: a very short introduction*, Vol. 75, Oxford Paperbacks, Oxford University Press, Oxford.
- Dahl, R. A. (1971), *Polyarchy: Participation and opposition*, Yale University Press, New Haven and London.
- Diamond, L. (1999), *Developing democracy: Toward consolidation*, Johns Hopkins University Press, Baltimore Maryland, USA.
- Eckstein, H. (1973), 'Authority patterns: A structural basis for political inquiry', *American Political Science Review* **67**(4), 1142–1161.
- Elff, M. & Ziaja, S. (2018), 'Method factors in democracy indicators', *Politics and Governance* **6**(1), 92–104.
- Elkins, Z. (2000), 'Gradations of democracy? empirical tests of alternative conceptualizations', *American Journal of Political Science* **44**(2), 293–300.
- Fearon, J. D. & Laitin, D. D. (2003), 'Ethnicity, insurgency, and civil war', *American political science review* **97**(1), 75–90.
- Franco, Á., Álvarez-Dardet, C. & Ruiz, M. T. (2004), 'Effect of democracy on health: ecological study', *Bmj* **329**(7480), 1421–1423.
- Gates, S., Hegre, H., Jones, M. P. & Strand, H. (2006), 'Institutional inconsistency and political instability: Polity duration, 1800–2000', *American Journal of Political Science* **50**(4), 893–908.
- Gleditsch, K. S. & Ward, M. D. (1997), 'Double take: A reexamination of democracy and autocracy in modern polities', *Journal of Conflict Resolution* **41**(3), 361–383.
- Goertz, G. (2006), *Social science concepts: A user's guide*, Princeton University Press, Princeton, New Jersey, USA.
- Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J. & Woodward, M. (2010), 'A global model for forecasting political instability', *American Journal of Political Science* **54**(1), 190–208.
- Gurr, T. R. (1974), 'Persistence and change in political systems, 1800–1971', *American Political Science Review* **68**(4), 1482–1504.
- Hegre, H. (2001), 'Toward a democratic civil peace? democracy, political change, and civil war, 1816–1992', *American political science review* **95**(1), 33–48.
- Högström, J. (2013), 'Does the choice of democracy measure matter? comparisons between the two leading democracy indices, freedom house and polity iv', *Government and Opposition* **48**(2), 201–221.



- House, F. (2017a), 'Freedom in the world 2016: Anxious dictators, wavering democracies: Global freedom under pressure', Available at: [https://freedomhouse.org/sites/default/files/FH\\_FITW\\_Report\\_2016.pdf](https://freedomhouse.org/sites/default/files/FH_FITW_Report_2016.pdf) (accessed on Dec. 1, 2017) .
- House, F. (2017b), 'Methodology website', Available at: <https://freedomhouse.org/report/freedom-world-2016/methodology> (accessed on Dec. 1, 2017) .
- Høyland, B., Moene, K. & Willumsen, F. (2012), 'The tyranny of international index rankings', *Journal of Development economics* **97**(1), 1–14.
- Huntington, S. P. (1993), *The third wave: Democratization in the late twentieth century*, Vol. 4, University of Oklahoma Press, Norman, USA.
- Inkeles, A. (1991), *On measuring democracy: Its consequences and concomitants*, Transaction Publishers, New Brunswick, New Jersey, USA.
- Knutsen, C. H. (2010), 'Measuring effective democracy', *International Political Science Review* **31**(2), 109–128.
- Knutsen, C. H. (2011), 'The economic effects of democracy and dictatorship', *PhD Thesis* .
- Lindberg, S. I., Coppedge, M., Gerring, J. & Teorell, J. (2014), 'V-dem: A new way to measure democracy', *Journal of Democracy* **25**(3), 159–169.
- Mainwaring, S., Brinks, D. & Pérez-Liñán, A. (2007), 'Classifying political regimes in latin america, 1945-2004', In *"Regimes and democracy in Latin America: Theories and methods"*. Edited by Gerardo L. Munck. Oxford University Press, Oxford, UK pp. 123–160.
- Marquardt, K. L. & Pemstein, D. (2018), 'Irt models for expert-coded panel data', *Political Analysis* **26**(4), 431–456.
- Marshall, M. G., Gurr, T. R. & Jaggers, K. (2017a), 'Polity iv project: Dataset users' manual - political regime characteristics and transitions, 1800-2016', Available at <http://www.systemicpeace.org/inscrdata.html> (accessed on July 6, 2017) .
- Marshall, M. G., Gurr, T. R. & Jaggers, K. (2017b), 'Polity iv project: Political regime characteristics and transitions, 1800-2016', Available at <http://www.systemicpeace.org/inscrdata.html> (accessed on July 6, 2017) .
- Munck, G. L. (2009), *Measuring democracy: A bridge between scholarship and politics*, Johns Hopkins University Press, Baltimore, Maryland.
- Munck, G. L. & Verkuilen, J. (2002), 'Conceptualizing and measuring democracy: Evaluating alternative indices', *Comparative political studies* **35**(1), 5–34.
- Pemstein, D., Meserve, S. A. & Melton, J. (2010), 'Democratic compromise: A latent variable analysis of ten measures of regime type', *Political Analysis* **18**(4), 426–449.
- Przeworski, A., Alvarez, M., Cheibub, J. A. & Limongi, F. (2000), *Democracy and development: political institutions and well-being in the world, 1950-1990*, Vol. 3, Cambridge University Press, Cambridge, UK.
- Schmitter, P. C. & Karl, T. L. (1991), 'What democracy is... and is not', *Journal of democracy* **2**(3), 75–88.
- Schumpeter, J. A. (1976), *Capitalism, socialism and democracy [1942]*, Routledge: London and New York.
- Treier, S. & Jackman, S. (2008), 'Democracy as a latent variable', *American Journal of Political Science* **52**(1), 201–217.

- Vanhanen, T. (2000), 'A new dataset for measuring democracy, 1810-1998', *Journal of Peace Research* **37**(2), 251-265.
- Vreeland, J. R. (2008), 'The effect of political regime on civil war: Unpacking anocracy', *Journal of Conflict Resolution* **52**(3), 401-425.

ONLINE APPENDIX TO "HOW (NOT) TO MEASURE DEMOCRACY"  
Vanessa A. Boese

**A Factor Analysis of the Polity2 Index components:**

Coppedge et al. (2008) carry out a factor analysis for four of the components (parcomp, xropen, xrcomp and xconst) as well as several other measures of democracy. They find evidence for two latent factors - contestation and participation. In their analysis xropen contributes to the inclusiveness/participation factor while the other components are associated with the contestation factor. The results of their analysis might be misleading insofar as Democ, Autoc and Polity Index are not made up of the actual values of the respective components. Due to the intricate weighing and aggregation scheme the indices are sums of the scores on each the components categories. Hence, I created score variables capturing the actual value which is contributed to the Polity Index and carried out an exploratory factor analysis on them. The results for the factor analysis using the score variables are displayed in Table 11. One latent factor explains 95.86% of the variation the Polity Index's components - contestation (or, as Munck and Verkuilen put it: the procedure by which the executive office is filled). These findings in combination with the absence of any suffrage/inclusiveness requirement suggest that the Polity Index in fact is a measure of political contestation rather than democracy - even if one embraces the minimalist democracy definition with contestation and participation.

Table 11: Factor analysis results for the score variables of Polity's five components

Factor	Eigenvalue	Proportion
1	3.56	0.9586
2	0.36	0.0981
3	-0.01	-0.0038
N=	17,228	

Variable: scores of...	Factor Loadings (Factor 1)	Uniqueness	KMO
...xrcomp	0.8969	0.1955	0.7771
...xropen	0.8337	0.3050	0.7778
...xconst	0.8039	0.3537	0.9209
...parcomp	0.8504	0.2768	0.7833
...parreg	0.8314	0.3087	0.7925
		Overall:	0.8055

## B Checklist of questions and respective scores for the components of the Freedom House Index

Table 12: Checklist for the Freedom House Index. Source: Freedom House Methodology Website, House (2017b)

Score	Political Rights	Score	Civil Liberties
0-12 points	<b>A. Electoral Process</b> <ol style="list-style-type: none"> <li>Is the head of government or other chief national authority elected through free and fair elections?</li> <li>Are the national legislative representatives elected through free and fair elections?</li> <li>Are the electoral laws and framework fair?</li> </ol>	0-16 points	<b>D. Freedom of Expression and Belief</b> <ol style="list-style-type: none"> <li>Are there free and independent media and other forms of cultural expression? (<i>Note:</i> In cases where the media are state controlled but offer pluralistic points of view, the survey gives the system credit.)</li> <li>Are religious institutions and communities free to practice their faith and express themselves in public and private?</li> <li>Is there academic freedom, and is the educational system free of extensive political indoctrination?</li> <li>Is there open and free private discussion?</li> </ol>
0-16 points	<b>B. Political Pluralism and Participation</b> <ol style="list-style-type: none"> <li>Do the people have the right to organize in different political parties or other competitive political groupings of their choice, and is the system open to the rise and fall of these competing parties or groupings?</li> <li>Is there a significant opposition vote and a realistic opportunity for the opposition to increase its support or gain power through elections?</li> <li>Are the people's political choices free from domination by the military, foreign powers, totalitarian parties, religious hierarchies, economic oligarchies, or any other powerful group?</li> <li>Do cultural, ethnic, religious, or other minority groups have full political rights and electoral opportunities?</li> </ol>	0-12 points	<b>E. Associational and Organizational Rights</b> <ol style="list-style-type: none"> <li>Is there freedom of assembly, demonstration, and open public discussion?</li> <li>Is there freedom for nongovernmental organizations? (<i>Note:</i> This includes civic organizations, interest groups, foundations, etc., with an emphasis on those engaged in human rights- and governance-related work.)</li> <li>Are there free trade unions and peasant organizations or equivalents, and is there effective collective bargaining? Are there free professional and other private organizations?</li> </ol>
0-12 points	<b>C. Functioning of Government</b> <ol style="list-style-type: none"> <li>Do the freely elected head of government and national legislative representatives determine the policies of the government?</li> <li>Is the government free from pervasive corruption?</li> <li>Is the government accountable to the electorate between elections, and does it operate with openness and transparency?</li> </ol>	0-16 points	<b>F. Rule of Law</b> <ol style="list-style-type: none"> <li>Is there an independent judiciary?</li> <li>Does the rule of law prevail in civil and criminal matters? Are police under direct civilian control?</li> <li>Is there protection from political terror, unjustified imprisonment, exile, or torture, whether by groups that support or oppose the system? Is there freedom from war and insurgencies?</li> <li>Do laws, policies, and practices guarantee equal treatment of various segments of the population?</li> </ol>
(0-4 points)	<b>Additional Discretionary Political Rights Questions</b> <ol style="list-style-type: none"> <li>For traditional monarchies that have no parties or electoral process, does the system provide for genuine, meaningful consultation with the people, encourage public discussion of policy choices, and allow the right to petition the ruler?</li> </ol>	0-16 points	<b>G. Personal Autonomy and Individual Rights</b> <ol style="list-style-type: none"> <li>Do individuals enjoy freedom of travel or choice of residence, employment, or institution of higher education?</li> <li>Do individuals have the right to own property and establish private businesses? Is private business activity unduly influenced by government officials, the security forces, political parties/organizations, or organized crime?</li> <li>Are there personal social freedoms, including gender equality, choice of marriage partners, and size of family?</li> <li>Is there equality of opportunity and the absence of economic exploitation?</li> </ol>
(-4 to 0 points)	<ol style="list-style-type: none"> <li>Is the government or occupying power deliberately changing the ethnic composition of a country or territory so as to destroy a culture or tip the political balance in favor of another group?</li> </ol> <p><b>Note:</b> For additional discretionary question A, a score of 1 to 4 may be added, as applicable, while for discretionary question B, a score of 1 to 4 may be subtracted, as applicable (the worse the situation, the more points may be subtracted).</p>		
Total	0 - 40 points	With the two discretionary questions the highest possible score remains 40, but the lowest possible score is -4.	0 - 60 points

## C Setup of V-Dem’s Electoral Democracy Index (Polyarchy)

Table 13: Setup of the Electoral Democracy (Polyarchy) Index. The information displayed here is gathered from the V-Dem Codebook, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Altman, Andersson, Bernhard, Fish, Glynn et al. (2017), p.49 - 59 and p. 435-436.

Electoral Democracy Index, **v2x\_polyarchy**,

*Question: To what extent are rulers responsive to citizens?(...) [It is] achieved through electoral competition for the electorate’s approval under circumstances when suffrage is extensive; political and civil society organizations can operate freely; elections are clean and not marred by fraud or systematic irregularities; and elections affect the composition of the chief executive of the country. In between elections, there is freedom of expression and an independent media capable of presenting alternative views on matters of political relevance.*

*The aggregation is done at the level of Dahl’s sub-components (with the one exception of the non-electoral component). The index is aggregated using this formula:*

$$v2x\_polyarchy = \frac{1}{2}v2x\_api + \frac{1}{2}v2x\_mpi$$

### Mid-Level Indices:

Additive Polyarchy Index, **v2x\_api**

Multiplicative polyarchy index, **v2x\_mpi**

*Question: To what extent is the electoral principle of democracy achieved?*

*Clarification: The electoral principle of democracy seeks to achieve responsiveness and accountability between leaders and citizens through the mechanism of competitive elections. This is presumed to be achieved when suffrage is extensive; political and civil society organizations can operate freely; elections are clean and not marred by fraud or systematic irregularities; and the chief executive of a country is selected (directly or indirectly) through elections.*

$$v2x\_api = \frac{1}{4}v2x\_frassoc\_thick + \frac{1}{4}v2xel\_frefair + \frac{1}{4}v2x\_freexp\_thick + \frac{1}{4}\left(\frac{1}{2}v2x\_elecoeff + \frac{1}{2}v2x\_suffr\right)$$

$$v2x\_mpi = v2x\_frassoc\_thick \cdot v2xel\_frefair \cdot v2x\_freexp\_thick \cdot \dots \cdot v2x\_elecoeff \cdot v2x\_suffr$$

### Low-Level Indices:

Freedom of expression index, <b>v2x_freexp_thick</b>	Freedom of association index, <b>v2x_frassoc_thick</b>	Clean elections index, <b>v2xel_frefair</b>	Elected official index, <b>v2x_elecoeff</b>	Share of population with suffrage, <b>v2x_suffr</b>
Type A (factual), B and C (expert coded) variables				
Government censorship effort of media, Harassment of journalists, Media self-censorship, Media bias, Print/broadcast media critical, Print/broadcast media perspectives; Freedom of discussion for men; Freedom of discussion for women; Freedom of academic and cultural expression)	Party ban, Barriers to parties, Opposition parties autonomy, Elections multiparty, civil society entry and exit, civil society repression	EMB autonomy, EMB capacity, Election voter registry, Election vote buying, Election other voting irregularities, Election government intimidation, Election other electoral violence, Election free and fair	Legislature bicameral, Lower chamber elected, Upper chamber elected, Legislature dominant chamber, head of state(HOS) selection by legislature in practice, HOS appointment in practice, HOG selection by legislature in practice, HOG appointment in practice, HOS appoints cabinet in practice, HOG appoints cabinet in practice, HOS dismisses ministers in practice, HOG dismisses ministers in practice, HOS = HOG? Chief executive appointment by upper chamber, Chief executive appointment by upper chamber explicit approval	Percent of population with suffrage

## D Derivation of the Polyarchy Index' rate of influence

$$\text{democracy index} = \frac{1}{4} \cdot (\text{polyarchy}^{1.6} + \text{component index}) + \frac{1}{2} \cdot \text{polyarchy}^{1.6} \cdot \text{component index} \quad (1)$$

The polyarchy index influences each high level democracy index to the power of 1.6. The intuition behind this rate is explained in the V-Dem Methodology V7 paper, Coppedge, Gerring, Lindberg, Skaaning, Teorell, Krusell, Marquardt, Mechkova, Pemstein, Pernes et al. (2017), p.10: "when a country has a polyarchy score of .5 (in practice, this is a threshold on the Electoral Democracy Index beyond which countries tend to be considered electoral democracies in a minimal sense) and its HPC<sup>113</sup> is at its maximum (1), the high level index score should be .5".

$$0.5 = \frac{1}{4} \cdot (0.5^x + 1) + \frac{1}{2} \cdot 0.5^x \cdot 1 \rightarrow x \approx 1.6 \quad (2)$$

This benchmark case is shown in equation 1. Solving for x yields a rate of close to 1.6. Intuitively, setting a higher rate of influence for polyarchy than any of the other component indices reflects a notion of democracy being a question of kind before one of degree. The principles of contestation and participation should be satisfied to a certain degree before further aspects of democracy can be employed to distinguish between regime types.

## E Summary statistics for the V-Dem democracy measures

Summary statistics for...						
Democracy Index:	...all observations available per index					
	Obs.	Mean	Median	St. D.	Min	Max
Polyarchy (Electoral)	17,036	0.318	0.206	0.279	0.009	0.947
Liberal	17,035	0.260	0.151	0.246	0.009	0.916
Participatory	17,035	0.192	0.105	0.193	0.005	0.814
Deliberative	17,035	0.209	0.068	0.262	0.000	0.913
Egalitarian	17,036	0.242	0.148	0.232	0.011	0.890

...the trunk dataset						
Index:	Obs.	Mean	Median	St. D.	Min	Max
Polyarchy (Electoral)	6,546	0.455	0.406	0.287	0.014	0.947
Liberal	6,546	0.354	0.260	0.279	0.010	0.916
Participatory	6,546	0.279	0.212	0.210	0.007	0.814
Deliberative	6,546	0.330	0.236	0.294	0.001	0.913
Egalitarian	6,546	0.356	0.260	0.249	0.032	0.890

Table 14: Summary Statistics for V-Dem democracy measures for all observations available (left) and observations in the trunk dataset (right).

<sup>113</sup>HPC refers to High Principle Component (here: component index).

## F Regression of difference between democracy measures on "hypothetical democracy"

Dep. Var.:	Polity2 - Polyarchy Coeff.	Polity2 - FHI Coeff	FHI-Polyarchy Coeff
hypothetical democracy	0.2326*** (0.0055)	0.0857*** (0.0064)	0.1469*** (.0049)
constant	-0.0040 (0.0033)	0.0120*** (0.0038)	-0.0241*** (.0029)
<i>R</i>	0.2175	0.0266	0.1252
N	6,546	6,546	6,546

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.10$

Table 15: Regression results for Regression of difference between democracy measures on "hypothetical democracy". Pooled OLS, standard deviation in parenthesis below.

Independent variable: "hypothetical democracy" :=  $\frac{(Polity2+FHI+Polyarchy)}{3}$

## G Description of the democracy measures' distribution

Figure 12: Histogram of the normalized democracy measures in trunk dataset

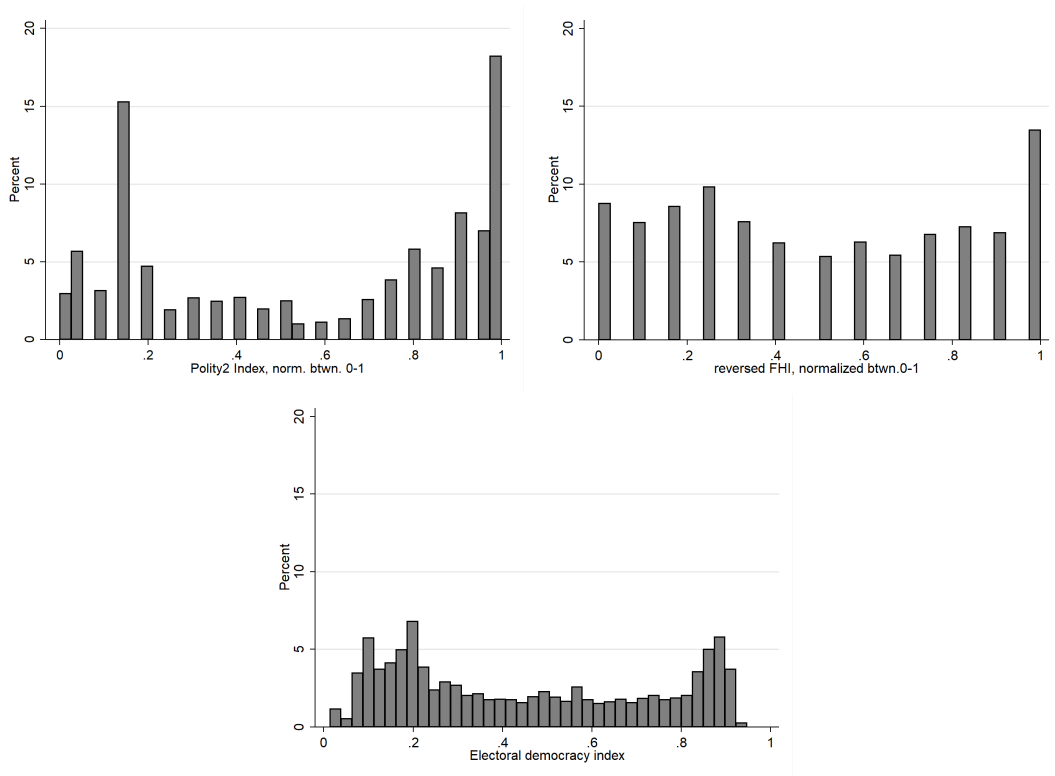
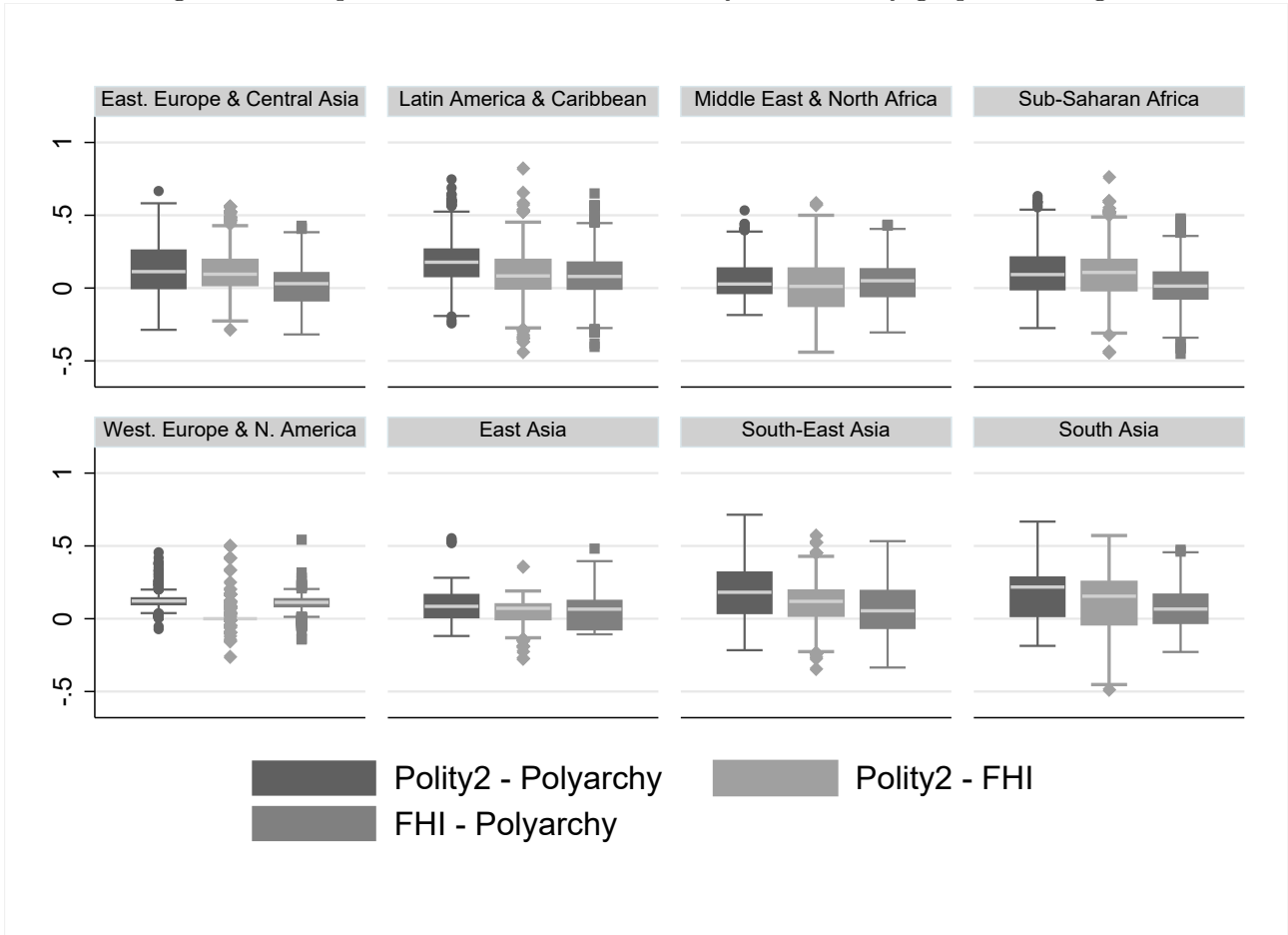


Table 16: Percentiles for Polity2 (normalized between 0-1), Polyarchy and FHI (reversed and normalized between 0-1) in trunk dataset

	Polity2		Polyarchy		FHI	
	Percentiles	Smallest	Percentiles	Smallest	Percentiles	Smallest
1%	0	0	0.0257	0.0140	0	0
5%	0.05	0	0.0858	0.0140	0	0
10%	0.1	0	0.1069	0.0140	0.0833	0
25%	0.15	0	0.1911	0.0140	0.2500	0
50%	0.7		0.4061		0.5	
		Largest		Largest		Largest
75%	0.95	1	0.7380	0.9335	0.8333	1
90%	1	1	0.8725	0.9357	1	1
95%	1	1	0.8912	0.9448	1	1
99%	1	1	0.9103	0.9471	1	1



Figure 13: Boxplot of differences in democracy measures by geopolitical region



## H Examining differences in regional codings and rankings

To examine the distribution of the differences between the index values a boxplot by region is provided in Figure 13. Polity2 mostly assigns higher values than the other indices, followed by the FHI. The inter quartile range of the differences between Polity2 and the other two indices is non-negative for all regions except MENA and South Asia (for Polity2 - FHI). There is a very high level of agreement between all indices for Western Europe and North America (the differences are mostly positive but close to 0). All indices also display a high level of agreement for East Asia. The picture for differences between FHI and Polyarchy is mixed, although the FHI mostly assigns larger values than the Polyarchy Index (median differences are positive for all regions, interquartile range non-negative for Latin America & the Caribbean, Western Europe & North America).

### H.0.1 Ranking

While the differences discussed above provide some information regarding the general "austerity" of each index they do not give any information on how the countries are coded in each year with respect to each other. When ranking all countries according to their democracy index value in a given year is each country ranked consistently across the three measures? In the following the country rankings will be compared across the three democracy measures. Note, that this exercise has its limitations due to the uncertainty embedded in each measure. For an excellent discussion of this see Høyland et al. (2012). For each year, all countries were ranked according to their democracy index values obtaining a rank in Polity2 values, a rank in FHI values and a rank in Polyarchy values. The regional average democracy values from Figure 5 are thus reproduced as regional average ranks in Figure 14. A rank of 1 corresponds to the lowest possible democracy index value. The ranks are coded keeping the overall

Figure 14: Regional average ranks based on the yearly values coded by each of the democracy indices

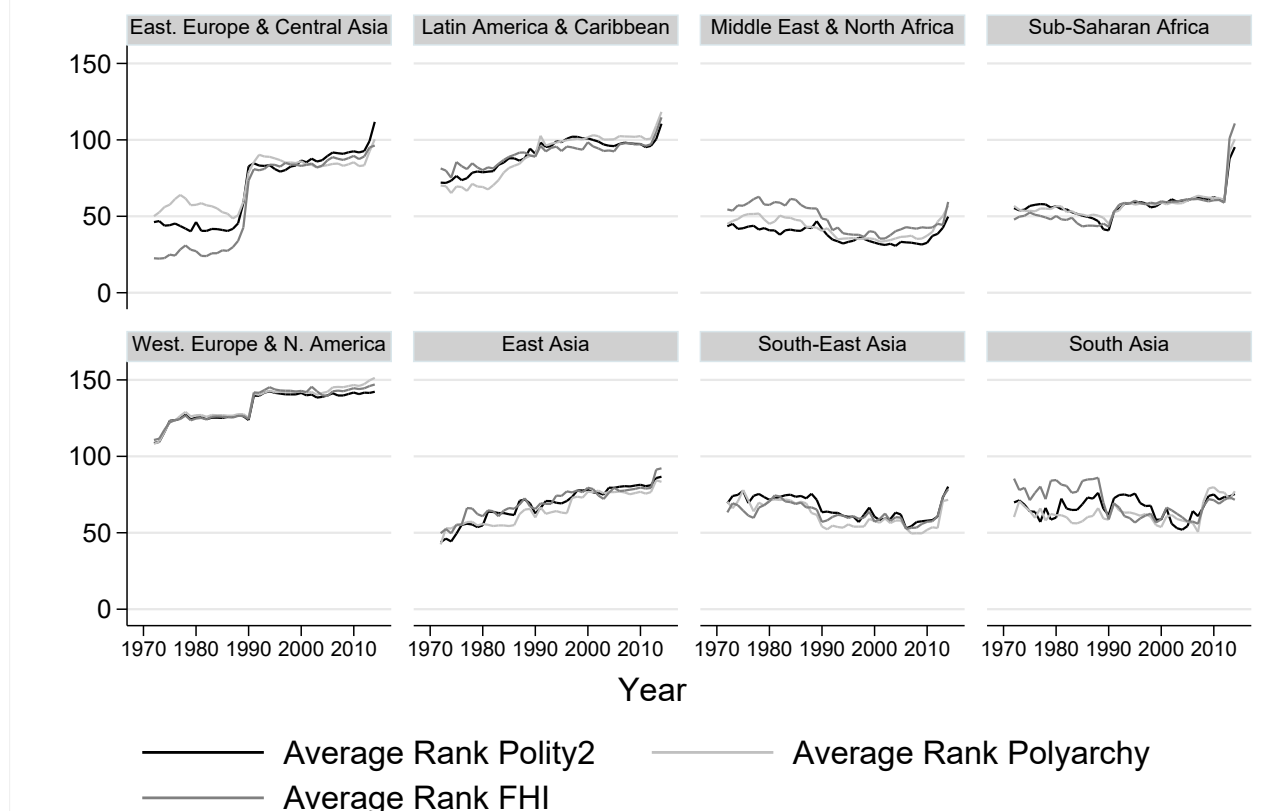


Table 17: Kendall's  $\tau_b$  (upper right side) and Spearman's  $\rho$  (bottom left side)

	Rank Polity2	Rank Polyarchy	Rank FHI
Rank Polity2	-	0.6903***	0.6909***
Rank Polyarchy	0.8735***	-	0.7303***
Rank FHI	0.8729***	0.8993***	-

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.10$

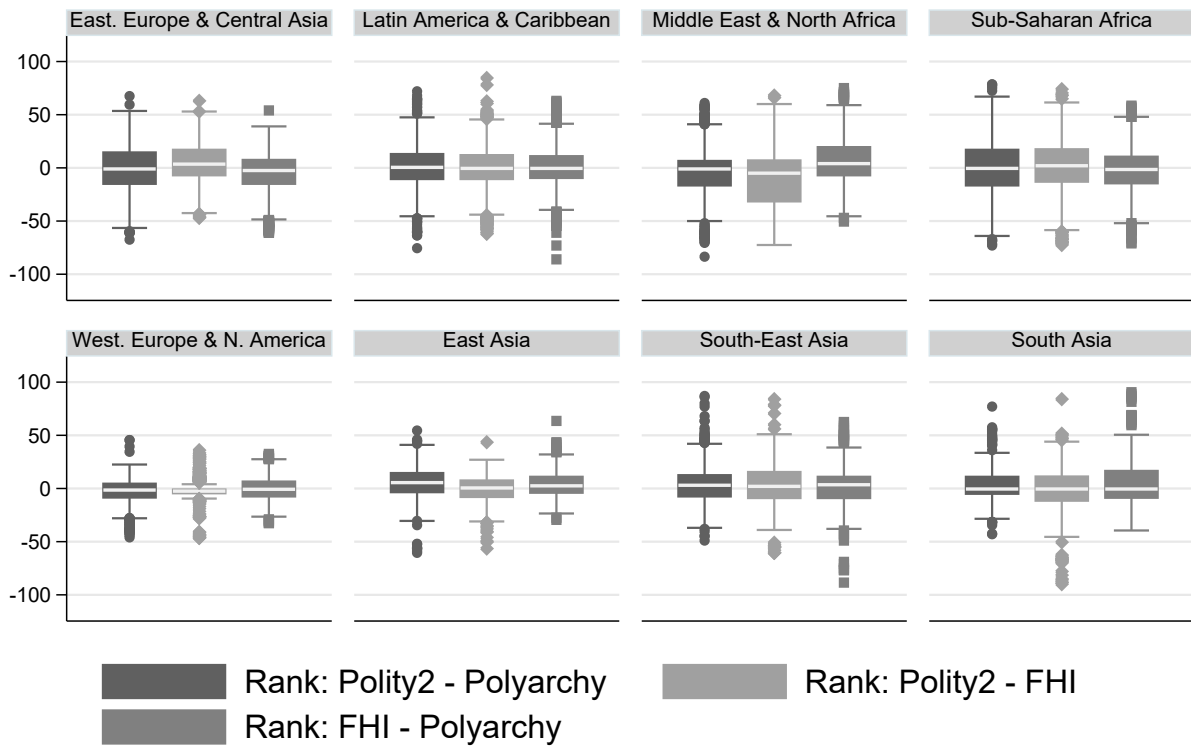
sum constant. Note that the ranking should only be compared within single years but not over time as one country's ranking can change when other countries are coded as democratic.

Western European & North American, Sub-Saharan African and East Asian countries are ranked very similarly by the indices. The difference in values for Eastern Europe & Central Asia translates into the most pronounced difference in rankings. Hence, in research frameworks in which countries of that time period and region are considered it is very likely that the results will vary with the index used. Therefore, using the FHI in such cases is inadvisable. Further notable differences in rankings occur before 1990 in the MENA region and South Asia. In these cases, FHI assigns higher average democracy values than the other two indices.

Figure 15 displays a boxplot of the differences between index rankings by region. The difference in ranks has the highest variability for Polity2-Polyarchy in Sub-Saharan Africa. Figure 15 shows that the medians for the difference in ranges are in most cases close to and the interquartile ranges centered around 0. This suggests that the differences in index values do not translate into difference in ranks. Lastly, the similarities of the rankings are compared using Kendall's  $\tau$ <sup>114</sup> and Spearman's  $\rho$ . A  $\tau$  or a  $\rho$  close to 1 implies a strong association between the rankings. Spearman's  $\rho$  "accepts" small differences in ranking from time to time and is very sensitive to large errors (even if they occur only

<sup>114</sup>To include ties we used  $\tau_b$ .

Figure 15: Boxplot of differences in yearly country rankings between indices by region

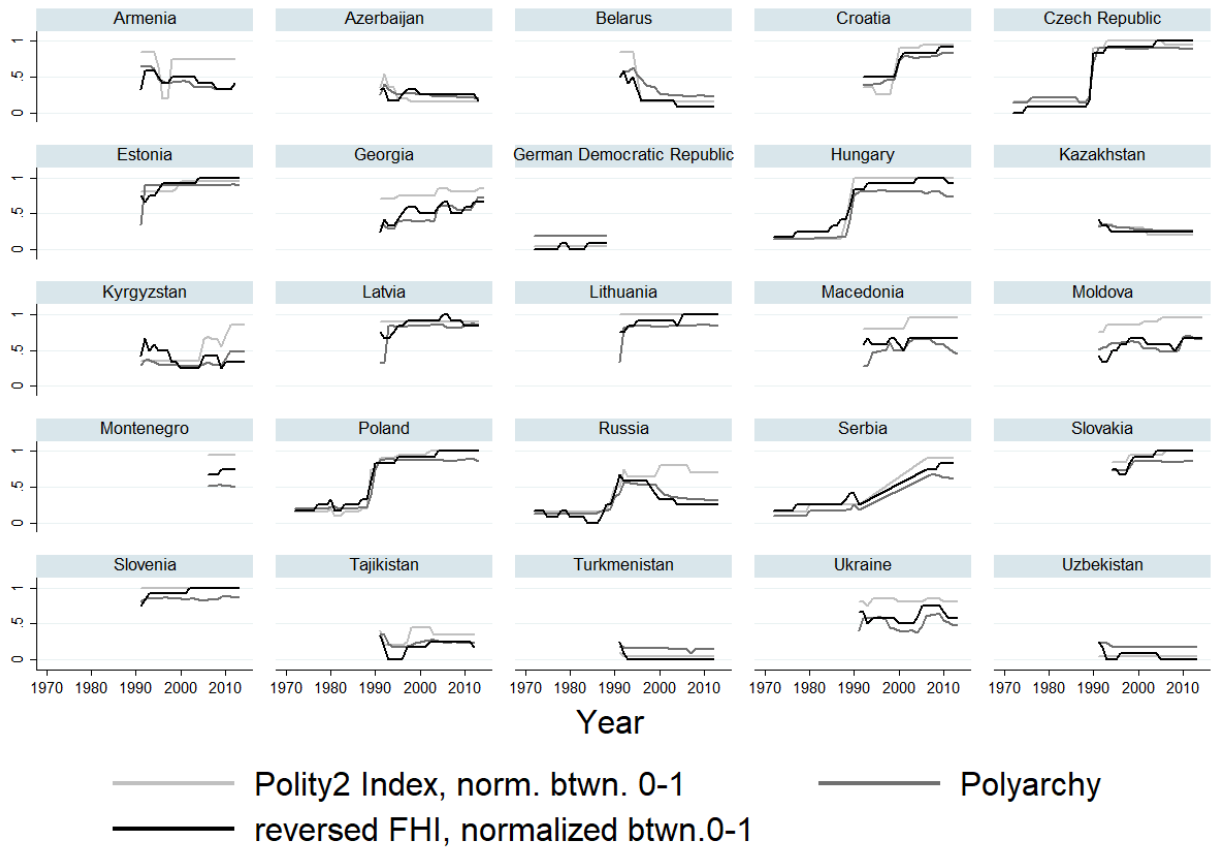


once). Kendall's  $\tau$  on the other hand is relatively insensitive to large errors occurring just once. The rank correlation coefficients are displayed in Table 17. Both, Kendall's  $\tau$  and Spearman's  $\rho$ , point to a very strong agreement with respect to the ranking for all variables. The nullhypothesis of independent rankings is rejected in all pairwise comparisons. In sum, the vast majority of country years is ranked consistently across all indices.

# I Eastern Europe and Central Asia (EECA)

Figure 16 displays the country codings by Polity2, FHI and Polyarchy for each country in the EECA region with more than 7 years of data available in the trunk dataset (Kosovo and Bosnia and Herzegovina did not meet this criterion and were thus left out, see Table 20).

Figure 16: Democracy coding by Polity2, FHI and Polyarchy for selected countries in Eastern Europe & Central Asia



## J Countries and their respective geopolitical regions

Table 18: Countries and their respective geopolitical regions as coded in section *Comparison*.

Region	Countries
1 East. Europe & Central Asia (31 countries)	Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Georgia, German Democratic Republic, Hungary, Kazakhstan, Kosovo, Kyrgyzstan, Latvia, Lithuania, Macedonia, Moldova, Mongolia, Montenegro, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine, Uzbekistan
2 Latin America & Caribbean (24 countries)	Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Suriname, Trinidad and Tobago, Uruguay, Venezuela
3 Middle East & North Africa (18 countries)	Algeria, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Saudi Arabia, South Yemen, Syria, Tunisia, Turkey, Yemen
4 Sub-Saharan Africa (47 countries)	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African, Chad, Comoros, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Republic of the Congo, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
5 West. Europe & North America (21 countries)	Australia, Austria, Belgium, Canada, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States
6 East Asia (5 countries)	China, Japan, North Korea, South Korea, Taiwan
7 South-East Asia (14 countries)	Burma/Myanmar, Cambodia, Democratic Republic of Vietnam, Fiji, Indonesia, Laos, Malaysia, Papua New Guinea, Philippines, Republic of Vietnam, Singapore, Solomon Islands, Thailand, Timor-Leste
8 South Asia (7 countries)	Afghanistan, Bangladesh, Bhutan, India, Nepal, Pakistan, Sri Lanka

## K Slow changing level of democracy in Polity2 Index

When using the Polity2 Index in time series and/or cross section models one factor to keep in mind is the slow changing nature of the level of democracy. In most years the level of a country's previous democracy index is the best predictor for its' current value. Gleditsch & Ward (1997) examined these changes in the Polity II Data (one of Polity IV's predecessors) with the help of Markov transition matrices. They show that much of the variation in the polity index is cross sectional rather than temporal. Table 19 examines the variation in the Polity2 Index more closely.

Table 19: Examining variation in the Polity2 Index

polity2	Overall		Between		Within
	Freq.	Percent	Freq.	Percent	Percent
-10	1354	7.97	33	17.01	42.54
-9	1144	6.73	62	31.96	23.75
-8	512	3.01	50	25.77	13.32
-7	1,884	11.09	107	55.15	25.44
-6	1297	7.63	82	42.27	13.80
-5	586	3.45	64	32.99	8.92
-4	715	4.21	65	33.51	12.41
-3	1,166	6.86	67	34.54	15.27
-2	324	1.91	48	24.74	9.03
-1	540	3.18	48	24.74	10.93
0	415	2.44	69	35.57	9.66
1	379	2.23	37	19.07	10.77
2	464	2.73	52	26.80	9.72
3	282	1.66	37	19.07	6.84
4	535	3.15	49	25.26	13.73
5	441	2.60	58	29.90	11.08
6	547	3.22	58	29.90	14.62
7	549	3.23	61	31.44	12.59
8	773	4.55	65	33.51	15.25
9	645	3.80	48	24.74	18.53
10	2,440	14.36	43	22.16	46.71
Total	16,992	100.00	1,203	620.10	16.13
			(n=194)		

For the overall part the unit of observation is a country-year. There are 1,354 observations in which a country in a given year obtained a polity index value of -10. In the between part the unit of observation is a country instead of a country-year; 33 of the countries ever had a Polity2 value of -10 and a total of 1203 countries was categorized. Due to the fact that the data only includes 194 countries, it follows that some countries switched between the categories. The within percent indicates the percentage of the time a country has the specified Polity2 value. Conditional on a country ever having a Polity2 value of -10, 42.54% of that same country's observations have the same index value. Interestingly, this percentage increases for both "high" democracies and autocracies. Conditional on a country ever having obtained a Polity2 value of 10, 46.71% of that country's observations have the same index value. This fraction is much smaller (around 10%) for Polity2 values between -2 and 2.

## L Countries and years included in the trunk dataset

Table 20: Countries (A-F) and years included in the trunk dataset

country	first year	last year	# years in sample	# of missing years
Afghanistan	1972	2015	21	23
Albania	1972	2015	44	0
Algeria	1972	2015	44	0
Angola	1975	2015	41	0
Argentina	1972	2015	44	0
Armenia	1991	2015	25	0
Australia	1972	2015	44	0
Austria	1972	2015	44	0
Azerbaijan	1991	2015	25	0
Bangladesh	1972	2015	44	0
Belarus	1991	2015	25	0
Belgium	1972	2015	44	0
Benin	1972	2015	44	0
Bhutan	1972	2015	44	0
Bolivia	1972	2015	44	0
Bosnia and Herzegovina	1993	1994	2	0
Botswana	1972	2015	44	0
Brazil	1972	2015	44	0
Bulgaria	1972	2015	44	0
Burkina Faso	1972	2015	44	0
Burma/Myanmar	1972	2014	43	0
Burundi	1972	2015	44	0
Cambodia	1972	2015	35	9
Cameroon	1972	2015	44	0
Canada	1972	2015	44	0
Cape Verde	1975	2015	41	0
Central African Republic	1972	2015	44	0
Chad	1972	2015	44	0
Chile	1972	2015	44	0
China	1972	2015	44	0
Colombia	1972	2015	44	0
Comoros	1975	2015	41	0
Costa Rica	1972	2015	44	0
Croatia	1991	2015	25	0
Cuba	1972	2015	44	0
Cyprus	1972	2015	44	0
Czech Republic	1972	2015	44	0
Democratic Republic of Congo	1972	2015	44	0
Democratic Republic of Vietnam	1972	2015	44	0
Denmark	1972	2015	44	0
Djibouti	1977	2015	39	0
Dominican Republic	1972	2015	44	0
Ecuador	1972	2015	44	0
Egypt	1972	2015	44	0
El Salvador	1972	2015	44	0
Equatorial Guinea	1972	2015	44	0
Eritrea	1993	2015	23	0
Estonia	1991	2015	25	0
Ethiopia	1972	2015	44	0
Fiji	1972	2015	44	0
Finland	1972	2015	44	0
France	1972	2015	44	0

Table 21: Countries (G-N) and years in trunk dataset continued

country	first year	last year	# years in sample	# of missing years
Gabon	1972	2015	44	0
Gambia	1972	2015	44	0
Georgia	1991	2015	25	0
German Democratic Republic	1972	1988	17	0
Germany	1972	2015	44	0
Ghana	1972	2015	44	0
Greece	1972	2015	44	0
Guatemala	1972	2015	44	0
Guinea	1972	2015	44	0
Guinea-Bissau	1974	2015	42	0
Guyana	1972	2015	44	0
Haiti	1972	2015	44	0
Honduras	1972	2015	44	0
Hungary	1972	2015	44	0
India	1972	2015	44	0
Indonesia	1972	2015	44	0
Iran	1972	2015	44	0
Iraq	1972	2015	37	7
Ireland	1972	2015	44	0
Israel	1972	2015	44	0
Italy	1972	2015	44	0
Ivory Coast	1972	2015	44	0
Jamaica	1972	2015	44	0
Japan	1972	2015	44	0
Jordan	1972	2015	44	0
Kazakhstan	1991	2015	25	0
Kenya	1972	2015	44	0
Kosovo	2009	2015	7	0
Kuwait	1972	2015	43	1
Kyrgyzstan	1991	2015	25	0
Laos	1972	2015	44	0
Latvia	1991	2015	25	0
Lebanon	1972	2015	29	15
Lesotho	1972	2015	44	0
Liberia	1972	2015	44	0
Libya	1972	2015	44	0
Lithuania	1991	2015	25	0
Macedonia	1992	2015	24	0
Madagascar	1972	2015	44	0
Malawi	1972	2015	44	0
Malaysia	1972	2015	44	0
Mali	1972	2015	44	0
Mauritania	1972	2015	44	0
Mauritius	1972	2015	44	0
Mexico	1972	2015	44	0
Moldova	1991	2015	25	0
Mongolia	1972	2015	44	0
Montenegro	2006	2015	10	0
Morocco	1972	2015	44	0
Mozambique	1978	2015	38	0
Namibia	1990	2015	26	0
Nepal	1972	2015	44	0
Netherlands	1972	2015	44	0
New Zealand	1972	2015	44	0
Nicaragua	1972	2015	44	0
Niger	1972	2015	44	0
Nigeria	1972	2015	44	0
North Korea	1972	2015	44	0
Norway	1972	2015	44	0



Table 22: Countries (O-Z) and years in trunk dataset continued

country	first year	last year	# years in sample	# of missing years
Oman	2000	2015	16	0
Pakistan	1972	2015	44	0
Panama	1972	2015	44	0
Papua New Guinea	1975	2015	41	0
Paraguay	1972	2015	44	0
Peru	1972	2015	44	0
Philippines	1972	2015	44	0
Poland	1972	2015	44	0
Portugal	1972	2015	44	0
Qatar	1972	2015	44	0
Republic of Vietnam	1973	1975	3	0
Republic of the Congo	1972	2015	44	0
Romania	1972	2015	44	0
Russia	1972	2015	44	0
Rwanda	1972	2015	44	0
Saudi Arabia	1972	2015	44	0
Senegal	1972	2015	44	0
Serbia	1972	2015	30	14
Sierra Leone	1972	2015	44	0
Singapore	1972	2015	44	0
Slovakia	1994	2015	22	0
Slovenia	1991	2015	25	0
Solomon Islands	1978	2015	37	1
Somalia	1972	2015	43	1
South Africa	1972	2015	44	0
South Korea	1972	2015	44	0
South Sudan	2012	2015	4	0
South Yemen	1972	1989	18	0
Spain	1972	2015	44	0
Sri Lanka	1972	2015	44	0
Sudan	1972	2011	40	0
Suriname	1975	2015	41	0
Swaziland	1972	2015	44	0
Sweden	1972	2015	44	0
Switzerland	1972	2015	44	0
Syria	1972	2015	44	0
Taiwan	1972	2015	44	0
Tajikistan	1991	2015	25	0
Tanzania	1972	2015	44	0
Thailand	1972	2015	44	0
Timor-Leste	2002	2015	14	0
Togo	1972	2015	44	0
Trinidad and Tobago	1972	2015	44	0
Tunisia	1972	2015	44	0
Turkey	1972	2015	44	0
Turkmenistan	1991	2015	25	0
Uganda	1972	2015	43	1
Ukraine	1991	2015	25	0
United Kingdom	1972	2015	44	0
United States	1972	2015	44	0
Uruguay	1972	2015	44	0
Uzbekistan	1991	2015	25	0
Venezuela	1972	2015	44	0
Yemen	1972	2015	43	1
Zambia	1972	2015	44	0
Zimbabwe	1972	2015	44	0