

Do democracy indices tell different stories?

Comparing eight democracy indices

Martin Paldam, Aarhus University ¹

Abstract:

The paper is an empirical comparison of eight democracy indices. It asks if they tell the same or diverse stories about the world. In large samples, the indices are highly correlated, as they have one strong common factor. It is the democratic transition, which is a long-run function of income. When projects compile more indices, they should measure something different, but the correlations of indices within projects are even higher than the correlations between indices from different projects. The C-scale converts the three main indices to the same percentage scale. It is used to compare the Polity and the Polyarchy indices. Their difference is an estimate of the gray zone of the index problem for democracy indices. It is at least 10 percentage points. The indices for 1/3 of the countries differ more than 10 pp, and they may even be uncorrelated. Thus, in small country samples it is important to replicate results with alternative indices.

Keywords: Democracy indices, democratic transition, aggregation problem,

Jel.: A12, K10, P51

Acknowledgements: The paper has received fine advice from Peter Nannestad. In addition, it reflects many discussions with Erich Gundlach. The paper was presented at the Silvaplana workshop 2021 and the SEA conference in Houston Texas, 2021. I thank the participants and others who have read the paper for many good comments.

¹ Department of Economics and Business, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.
E-mail mpaldam@econ.au.dk. Home page <http://www.martin.paldam.dk>.

1. Introduction

Comparative studies often consider a political dimension. Here a common tool is a democracy index that classifies political systems. About 15 democracy indices exist. Eight indices are issued by organizations, and hence they are regularly updated. They are compared in this paper. They can either be compared from the input side or from the output side.

The input side is covered by about 35 papers that discuss the indices conceptually and as regards their construction. Some references are Munck and Verkuilen (2002) who stress that the indices are made for different purposes, aggregate different indicators, etc. This is also the argument of Boese (2019), while Gründler and Krieger (2022) discuss the indices from the point of view of aggregation. This paper looks at the other side as a systematic comparison of output is missing.

Thus, the paper asks if the indices tell the same or alternative stories about the world. That is, if a paper uses one index to tell a certain story, can you trust that the other seven indices tell the same story, or is it necessary to replicate everything with the other indices? The paper has three main results:

(1) The indices are highly correlated in wide cross-country datasets due to one strong factor, which also includes *Income*.² The conceptual differences are dominated by the common factor. It is the democratic transition. It does not occur in the OPEC countries.

(2) The transition is a long-run causal relation from income to the political regime.

(3) When *Pol* and *Vpol* are converted to the same scale, they differ numerically by 12.4 pp (percentage points). At least 10 pp of this difference is the gray zone. It is not a measurement error, but imprecision due to the index problem.

Furthermore, while indices are highly correlated between projects, they are even more correlated within projects, where the indices are made to be different. Thus, conceptual differences between indices are dominated by the methods used in their compilation.

The indices assume two *anchors*: A top anchor for full/perfect democracy, and a bottom anchor for full/perfect authoritarianism. All countries are placed in the range between the anchors. The scales are assumed to be linear, though two of the scales use integers. There is fine agreement about the countries at the top, and some agreement about the countries at the bottom;³ see Figures 1 and 2 below and the Net-Appendix. As the indices are linear and use

² And so do the indices by Gründler and Krieger (2021), and by Bjørnskov and Rode (2020); see Gundlach (2021).

³ Most western countries are at the top, while the bottom are one-party dictatorships (as North Korea) and absolute kingdoms (as Saudi Arabia).

roughly the same anchors, they can be converted to the same *C-scale* as discussed in section 2.3. It is in percent of the range and has the same average.

Section 2 also discusses three general problems with the indices: (a) The *gray zone* of measurement uncertainty due to the index problem, (b) the weak *equidistance*, and (c) the two definitions of the *top anchor*. Section 3 compares the frequency distribution of the indices, while section 4 is a correlation/factor analysis showing the strength of the common factor. Section 5 display the common factor as a function of income. It is a perfect transition curve. Section 6 looks at the countries with the largest deviation between the indices. Finally, section 7 concludes. A Net-Appendix provides extra tables, graphs, and robustness tests.

Table 1 define the variables. The reader need not learn the table by heart, but it may be useful to tag it for easy reference. The paper concentrates on the main index from each project: *FHr*, *Pol*, and *Vpol* that become *FHrc*, *Polc*, and *Vpolc*, when converted to the same scale.

Table 1. Income and the eight democracy indices

Project	Index	Scale, with range and step width in % of range
Freedom House <i>FH</i>	(1) $CLr = 7 - CL$, Civil Liberties (2) $PRr = 7 - PR$, Political rights $FHr = (CLr + PRr)/2$ and $FHrc$ in the C-scale	Closed set of [7, 1] integers. 7 is fully authoritarian, 1 is fully democratic. When r is added to name it is rescaled to [0, 6]. One year is missing and has been interpolated. Step width is 16.7% for CL and PR , and 8.3% for FH
Polity	(3) Pol (Polity2) and $Polc$ in the C-scale	Closed set of [-10, 10] integers. -10 is fully authoritarian, 10 is fully democratic, 0 is no system. Step width is 5%
V-Dem	(4) $Vpol$, Polyarchy and $Vpolc$ in the C-scale (5) $Vlib$, liberal dem. (6) $Vpar$, participatory dem. (7) $Vdel$, deliberate dem. (8) $Vega$, egalitarian dem.	Open interval]0, 1[2-3 decimals. 0 is perfect authoritarian, 1 is perfect democracy. These ideals are not reached. The de facto width for $Vpol$ is [0.012, 0.926]. Step width is 0.1% though the index is often given with 1 decimal less, so the step width is 1%
Maddison	(9) $Income\ y = \ln\ gdp$	gdp is the real GDP per capita in PPP prices
Samples: (i) <i>All 155 countries</i> , $N = 6,599$, (ii) <i>Main</i> : 139 countries, $N = 5,872$, (iii) <i>OPEC</i> : 16 countries, $N = 727$		
The <i>C-scale</i> is defined in Table 2. It gives the indices the same average and the range 100. Thus, it is in pp (percentage points). It is used on the main indices <i>FHr</i> , <i>Pol</i> , and <i>Vpol</i> that become <i>FHrc</i> , <i>Polc</i> , and <i>Vpolc</i> .		



The paper uses all observations that are available across all series, 1972-2018. It is downloaded in July 2022. Observations where *Pol* is zero are excluded for all series. The references give the manuals for the indices and their home pages. Seven of the 155 countries are successors of an old country: The Czech Republic was Czechoslovakia; Germany was West Germany; Russia was USSR; Serbia was Yugoslavia; Sudan continues after South Sudan leaves; Vietnam was North Vietnam; and Yemen was Yemen North. In these cases, the data are joined into one series. **Terminology**: *Political system* and *regime* are used as synonyms. *Av* is the arithmetic average, and *std* is the standard deviation. Dem is democracy.

2. Theory

Section 2.1 introduces transition theory. Section 2.2 discusses the index problem for democracy indices, while section 2.3 reports a conversion method giving all indices the same scale. Finally, section 2.4 looks at the two choices of the top anchor – *full* vs *perfect* democracy.

2.1 Transition theory ⁴



Seen in a grand development perspective, two basic steady states exist: A *traditional* and a *modern*. The LICs (Low Income Countries) are close to the traditional steady state, while the HICs (High Income Countries) are in the modern steady state.

All socioeconomic series have different levels in the two steady states. They *diverge* from the traditional steady state when development starts, and much later they *converge* to the modern steady state.⁵ When a series is a ratio, the transition curve looks as  or , depending on the scaling of the series.⁶ In wide cross-country samples or long time series, the correlation between the series and income is 0.5 to 0.9. Transitions are an important, but fuzzy, underlying part of the dynamics of development. They generate much confluence in the series, so that all series with a transition can ‘explain’ each other, making causality elusive. Very resource rich countries get wealthy before the grand transition occurs, and hence has a different development. Thus, the data from past and present OPEC countries as in a separate *OPEC sample*, while the data from all other countries are the *Main sample*.

Kernel regression on a large, *unified* dataset makes transitions visible. The data can be either wide cross-country data samples or long time series. Typically, wide cross-country samples are available for a short period, and long time series for a small cross-country sample. They normally show much the same picture, confirming *equivalence* of the two dimensions as regards transitions. If the transition process generalizes and the data-samples are large, the kernel regression obtains narrow confidence limits. The kernel technique makes few assumptions on the functional form, and if the kernel obtains the form predicted, it is a strong test of transition theory. Section 5 shows that the democracy indices provide perfect transition curves though only for the Main data sample.

⁴ This subsection is based on Paldam (2021).

⁵ Standard texts on growth make a big point about the convergence to the modern steady state, while the divergence of growth from the traditional steady state has received much less attention; see Barro and Sala-i-Martin (2004) or Jones and Vollrath (2013). The divergence/convergence story also applies to provinces within countries.

⁶ The two shapes shown are in levels. Hence, the first difference is hump shaped  or .

2.2 *The index problem: A gray zone of equally good indices*

An index is an aggregate of a set of indicators. The aggregation process is unique in rare cases only. Thus, the index is surrounded by a gray zone of measurement uncertainty, which is the zone of equally good indices. It adds to the uncertainty that indices use different subsets of indicators and score the chosen indicators differently.

The price index aggregates prices that are conceptually simple and easy to measure. A literature has proved that the perfect price index does not exist.⁷ The reader may recall that the gray zone is the gap between the Laspeyres and the Paasche indices, where any average may be chosen. The zone is small in the short run, so useful price indices exist. The gray zone can be reduced by decreasing the interval between the polling of the baskets, but it should include the adjustment to the price changes, so the interval should not be too small.

Arrow (1963) proves that no political system can make a perfect aggregation of preferences, but there is surely a range of good institutions that can be combined – each has advantages and disadvantages, so compromises must be made. It is no wonder that regimes differ, even if they strive to be democratic. Thus, good democracy is possible, but perfect democracy is impossible.⁸

A democracy index aggregates diverse, often qualitative, and assessed indicators. Both the choice of indicators, the assessments and the aggregation have many variants. Here the gray zone of measurement uncertainty must be (much) larger than for the price index.⁹ Even if one measure is chosen, a range of equally good measures is possible, giving a (wide) gray zone.

The paper estimates the width of the gray zone by assuming that *Pol* (Polity) and *Vpol* (Polyarchy from V-Dem) try to measure almost the same. Both indices are made by a highly competent group, which has put years of work in its index. Thus, when they are rescaled to the same scale, as discussed in the next section, almost the full average numerical difference is a measure of the gray zone. Section 4.5 finds a gray zone of at least 10% (of the range).

2.3 *The C-scale and the equidistance problem*

Each project has its own scale, but they all report that they use linear scales. Two of the projects

⁷ The proofs follow Arrow (1963) by stating a set of conditions the perfect price index should have, and then show a contradiction; see Eichhorn and Voeller (1976). In practice, price indices require a choice of price set, a sampling plan, methods to deal with outliers, and the choice of aggregation formula, etc.

⁸ The author has participated in a detailed comparison of the remarkably different political systems in Denmark and Switzerland; see Christoffersen et al. (2014). The World Values Surveys show that the populations in the two countries think that their regime is fully democratic, as is also confirmed by the *FHR* and *Pol* indices.

⁹ Gründler and Krieger (2022) is a study of the aggregation problem using democracy indices as the illustration. They also show that the problem is substantial.

report their indices in integers. This makes sense given the large gray zone. The indices are converted to the same C-scale in Table 2 by two steps: (1) The range adjustment sets the distance between the two anchors, to 100, so that scores are in pp (percentage points). If the deviations from equidistance are stochastic, it means that once the indices are range-adjusted they should differ only by a constant. (2) The level adjustment uses the average difference as that constant. When the indices are in the C-scale, they are termed *adjusted*. The three adjusted series *FHrc*, *Polc* and *Vpolc* have the same range and the same average, and they are (almost) in percentage points.

Table 2. The C-scale adjusting to range and level, made for the Main sample

Index	Range	The two adjustments for the C-scale		Original Av (std)	C-scaled Av (std)
		Range adj.	Level adj.		
<i>FHr</i>	0 to 6	$FHc = 100(FHr-1)/6$	+6.680	4.29 (2.02)	61.55 (33.7)
<i>Pol</i>	-10 to 10	$Polc = 100(Pol + 10)/20$		2.31 (7.24)	61.55 (36.7)
<i>Vpol</i>	0 to 1	$Vpolc = 100Vpol$	+13.667	0.479 (0.29)	61.55 (29.0)

Av is the arithmetic average, i.e., the sum of the observations divided by their number. The C-index is almost in pp (percentage points), but thanks to the diverse top anchors (see sections 2.4 and 3.2) a small fraction of the *FHrc* and *Vpolc* scores are slightly over 100. The paper disregards this small problem.

The *anchors* are reasonably well defined, but the scoring between the anchors is somewhat arbitrary. Ideally, such scorings should be *equidistant*, but it is hard to see how this make sense. When one of the adjusted indices reports that country A scores x and country B scores $2x$, it says that there is twice as much democracy in B. Such ratios replicate poorly both across countries and over time. This is partly due to the gray zone. The largest level adjustment is between *Pol* and *Vpol*. Later Figures 3a and 5 will show that the systematic deviations are small for *Polc* and *Vpolc* as discussed in section 4.5.

Table 3. The indices for three years for Russia, using the C-scale

Years	Original scales			C-scale		
	<i>FHr</i>	<i>Pol</i>	<i>Vpol</i>	<i>FHrc</i>	<i>Polc</i>	<i>Vpolc</i>
1987	1.5	-7	0.146	15.0	15.0	28.3
1992	4.5	5	0.504	65.0	75.0	64.1
2018	1.4	4	0.272	15.0	70.0	40.9

Table 3 illustrates the equidistance problem. It looks at the scoring of Russia where the regime is between the anchors and in some flux. In addition, it is an important country, where

the projects must have made an extra effort. The indices agree that there was an important step towards democracy from 1987 to 1992, and that there has been a backslide since then. The table shows the three indices for the three relevant years. Equidistance is surely a problem.

2.4 *The top anchor problem (1): Full vs perfect democracy*

FHr and *Pol* are reported as integers; see Table 1. In the C-scale, the *FHrc* score has a step width of 8.3 pp, and *Polc* has a step width of 5 pp, both of which are smaller than the gray zone. Thus, **full** democracy is an interval that is attainable for both indices. However, even when the interval is smaller for the *Polc* index, more countries are in the top interval.

The top score for the *Vpol* index is 0.926 (in Sweden), indicating that democracy is **imperfect** even in the most democratic country. The lowest score is 0.012 (in Saudi Arabia); that is close to the theoretical minimum. Thus, the *Vpol* index has an open de facto range of]0.926, 0.012[and a closed theoretical range of [0, 1]. The C-scale uses the latter for the range adjustments to make it robust to future index scores.

Consequently, the *FH* and the *Pol* indices see democracy as a concrete set of legal rules that can be implemented in practice. Hence, they score (too?) many countries as full democracies. The V-Dem indices score democratic countries with a range that seems large compared to the differences at the lower part of the scale, pointing to the equidistance problem.

The top anchor problem is further analyzed in section 3.2, which looks at the *FHr* and *Vpol* scores for the 1,357 observations, where *Pol* is 10. See also the top of the distributions on Figure 1 below.

3 The distributions of the indices and the top-anchor difference

Section 3.1 compares the frequency distribution of the indices, while section 3.2 returns to the top anchor problem from the last section.

3.1 *The frequency distribution of the three main indices*

The number of bins used to count the frequencies is the same 20 for *Pol* and *Vpol*, but the scale of the *FH* index only allows 13 bins. All three distributions are low in the middle and have a strong peak at the top, and a weaker peak at the bottom. The *FH* index is symmetrical in the sense that the median and the mean are the same. The median of the *Pol* index is larger than the mean, so more countries are in the democratic range. The median of the *Vpol* index is

smaller than the mean. Thus, more countries are in the authoritarian range. The OPEC/MENA countries deviate substantially from the Main sample. Even when these countries have middle to high income, they have little democracy.

Figure 1. The frequency distribution of the three indices

Figure 1a
FHr index
(*FH* rescaled)

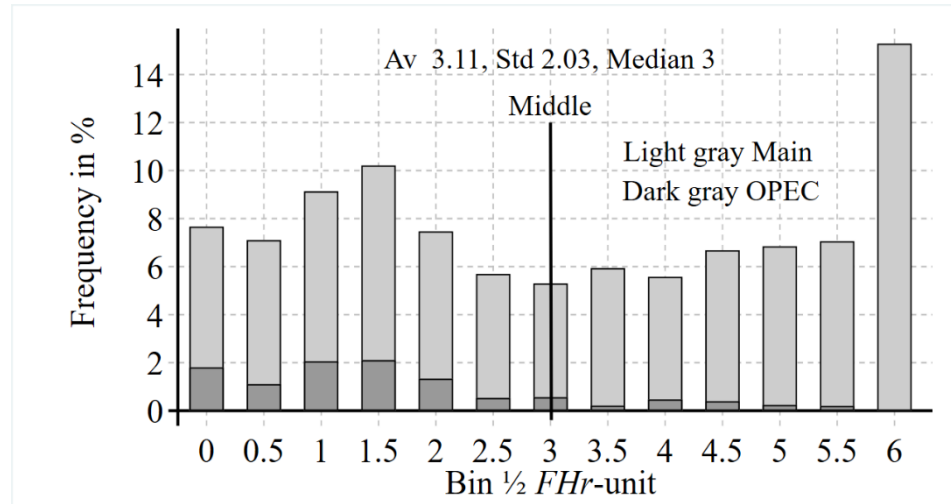


Figure 1b
Pol index
(Polity2)

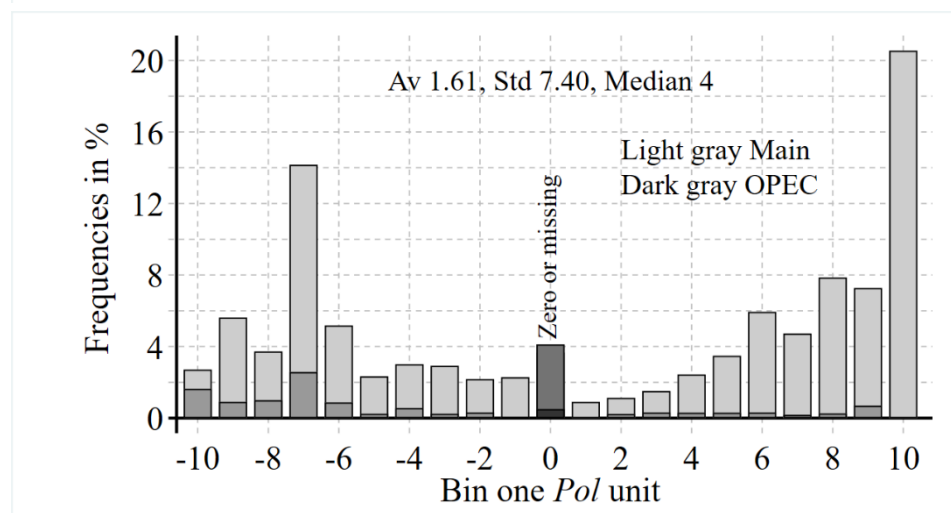
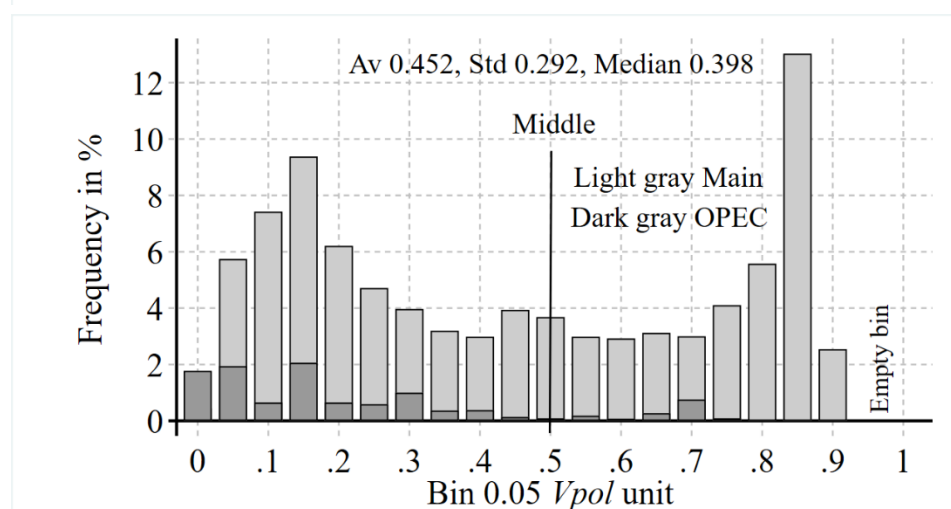


Figure 1c
Vpol index
(Polyarchy)



3.2 The top anchor problem (2): The observations where *Pol* is 10 for full democracy

The top anchor problem from section 2.4 is now reconsidered empirically. The *Pol* index scores 20% of the countries at 10 for full democracy. The *FH* index scores 15% of the countries as full democracies. The *Vpol* index has no observations in the highest bin. The other four V-Dem indices are even lower for the highest value observed. Still, Figure 1c does have a democracy peak for index values from 0.85 to 0.9.

Figure 2. The 1,357 observations where *Pol* is 10 for full democracy

Figure 2a. For the *FHr* index

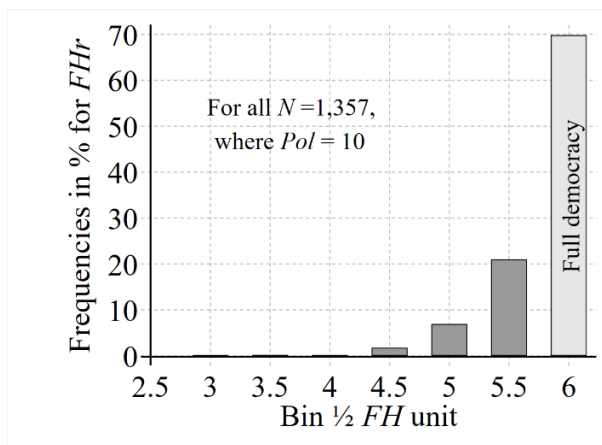


Figure 2b. For the *Vpol* index

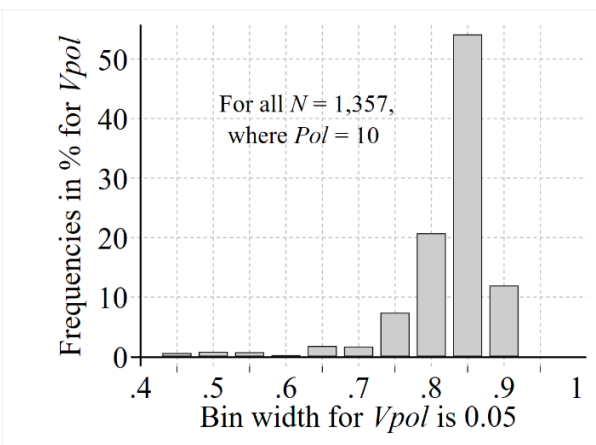


Figure 2c. The *income* dependency of the *Vpol* when *Pol* is 10

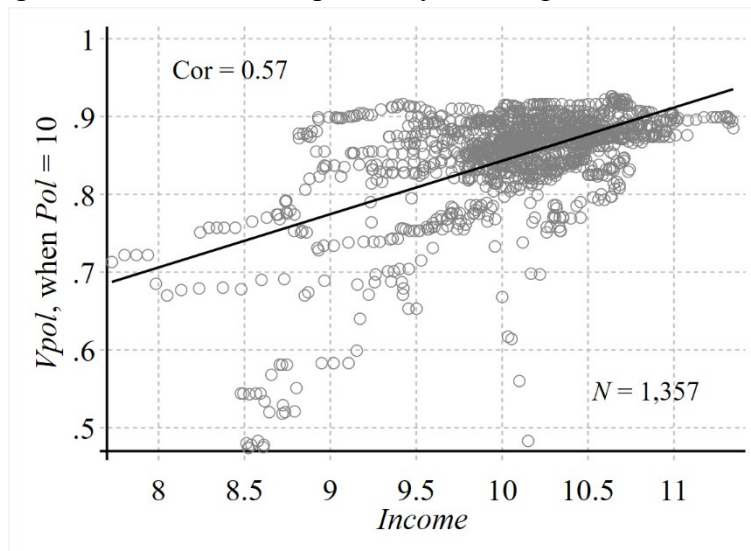


Figure 2 shows the frequency distribution of *FHr* and *Vpol* for the 1,357 observations where *Pol* is 10. The *FHr*-scores agree by having 6 for full democracy in 70% of the cases. Figure 2c further analyzes the 1,357 observations for *Vpol* data from Figure 2b. The largest

difference between *Pol* and *Vpol* occurs for middle-income countries. While *Pol* scores some such countries as full democracies, the V-Dem project is reluctant to accept this possibility. The two top anchor definitions can be interpreted in two ways:

(i) The *truncation* view: It sees *Pol* and *FH* as upward truncated, so that the top score hides something important. In this interpretation, the V-Dem project has de-truncated the scale.

(ii) The *equidistance* view: Differences at the top are small compared to the large ones at the lower end of the scale. Thus, the demand of equidistance makes the top collapse into one value. When the V-Dem indices increase the top-end of the scale, the lower part of the scale is correspondingly compressed.

4. Correlation and factor analysis

Section 4.1 looks at the closely related factor and correlation analysis, while section 4.2 compares the within-project correlations and the between-project correlations. Section 4.3 analyzes the importance of the anchors for the correlations.

4.1 Factor and correlation analysis

The factor analysis is reported in Table 4.

Table 4. Four factor analyses, 1972-2016

Factor	I. Main sample				II. OPEC sample			
	A1. Annual data <i>N</i> = 5,872		B1. Country av <i>N</i> = 139		A2. Annual data <i>N</i> = 727		B2. Country av <i>N</i> = 16	
	Eigenv	Cumul	Eigenv	Cumul	Eigenv	Cumul	Eigenv	Cumul
Factor1	7.84	0.97	8.14	0.97	6.89	0.92	7.33	0.86
Factor2	0.25	1.00	0.22	1.00	0.53	0.99	1.05	0.98
Variable	Factor loadings		Factor loadings		Factor loadings		Factor loadings	
	Factor1	Factor2	Factor1	Factor2	Factor1	Factor2	Factor1	Factor2
(1) <i>CLr</i>	0.93	-0.11	0.98	-0.10	0.81	-0.14	0.91	0.23
(2) <i>PRr</i>	0.94	-0.20	0.97	-0.17	0.90	-0.12	0.97	0.20
(3) <i>Pol</i>	0.90	-0.28	0.93	-0.28	0.90	-0.24	0.93	-0.31
(4) <i>Vpol</i>	0.99	-0.03	0.99	-0.02	0.98	-0.15	0.98	-0.16
(5) <i>Vlib</i>	0.99	0.09	0.99	0.09	0.97	0.11	0.98	0.11
(6) <i>Vpar</i>	0.98	0.06	0.98	0.04	0.96	-0.18	0.96	-0.15
(7) <i>Vdel</i>	0.98	0.05	0.98	0.06	0.96	0.11	0.97	0.11
(8) <i>Vega</i>	0.96	0.22	0.97	0.20	0.94	0.12	0.94	0.14
(9) <i>Income</i>	0.69	0.25	0.74	0.22	-0.02	0.58	-0.18	0.88

Eigenv means eigenvalue, and *Cumul* means cumulative. The gray shading indicates unreliable results. Factor2 in the last analysis is a borderline case where *N* = 16 only.

- (1) The four samples contain one and only one factor, i.e., the second and higher factors are negligible. The factor loading of all eight indices to the first factor are all very high.
- (2) The loading on income in the main sample is large and positive, while the loading on income is small and negative in the OPEC sample.

The Net-Appendix reports six tables of correlations, which are summarized in Table 5. The A and B columns are for the unified data and the country averages as in Table 4, while the C column is the average within-country correlation. It makes no sense for the factor analysis.

Table 5. Comparing between-project and within-project correlations

Correlations	<i>N</i>	I. Main sample			II. OPEC sample			
		A1. Unified	B1. Between	C1. Within	A2. Unified	B2. Between	C2. Within	
(1) All	Av	28	0.916	0.946	0.698	0.854	0.904	0.618
(2)	(2se)		(0.018)	(0.013)	(0.061)	(0.028)	(0.019)	(0.092)
(3) Between-projects	Av	17	0.883	0.926	0.589	0.811	0.886	0.494
(4)	(2se.)		(0.011)	(0.012)	(0.033)	(0.023)	(0.021)	(0.088)
(5)	Av+2se		0.894	0.938	0.622	0.834	0.907	0.583
(6) Within <i>FH</i>	One obs.	1	0.932	0.980	0.597	0.983	0.949	0.580
(7) Within V-Dem	Av	10	0.971	0.977	0.891	0.927	0.929	0.831
(8)	(2se)		(0.004)	(0.004)	(0.012)	(0.009)	(0.011)	(0.018)
(9) Test (2) = (5)	t-test, %		1.1/10 ¹⁰	7.6/10 ⁶	1.3/10 ¹¹	1.5/10 ⁷	1.8/10 ²	2.7/10 ⁵
(10) To income	Av	8	0.656	0.715	0.341	-0.016	-0.155	0.411

Rows (1) to (6) compares the $8 \times 7/2 = 28$ correlations between the indices. Row (6) compares the 17 between-project correlations and the 10 within-project correlations for the V-Dem indices.

The nine first rows deal with the correlations between the eight democracy-indices. Note first that (B) the results between countries are higher than (C) the result within countries. The result for (A) the unified data is in between. Thus, the correlation is larger in the long run than in the short run. This tally well with the finding in section 5, where all indices have a strong transition.

Row (1) of correlations to *Income* differ in panels (I) and (II). In the Main sample (I), the correlation between the democracy indices and *income* is always positive. It is strongest in the long run, but it is already substantial in the unified data; see sections 3.3 and 4. In the OPEC sample (II), most correlations are much smaller and negative in sections (A) and (B). However, here the within-countries correlations (C) are much the same as in the Main sample. Thus, the negative results are due to the high level of authoritarianism in the richest oil countries (Kuwait, Qatar, Saudi Arabia, and the UAE).

The results in sections (A) and (B) have a small range. This suggests that if the time unit is larger than a year (such as 5 years or 10 years), the result would stay in the range. This is indeed the case.

4.2 Within-project indices are more correlated: The fact and an interpretation

Row (3) in Table 5 reports the averages of the 17 correlations between projects. They are always smaller than the correlations in row (1). Row (5) shows their upper 95% confidence limit. Row (6) reports the correlations within the *FH*-project. They are all above or at the 95% limit in row (5), and so are all the 10 correlations within the V-Dem project. Row (9) gives the probability that the within-V-Dem project correlations and the between-projects correlations are the same. It is strongly rejected.

Thus, the within-project correlations are higher than the between-project correlations. The indices may differ for two reasons: (i) ***Because they measure something different.*** (ii) ***Because they measure it differently.*** Each project uses its own assessment package, i.e., a set of principles for the choice, scoring and aggregation of the indicators. Thus (ii) is more important than (i). This tallies with the high factor loadings on the within-project indices reported in the next section. Thus, the conceptual differences between democracy indices are of secondary importance. The devil is in the details, not in the concepts.

The Net-Appendix repeat the analysis above for the MENA countries, and find the same results, though marginally weaker. They also study the lead-lag structure between the democracy indices and show that neither index leads nor lags any other index significantly, though there is a weak tendency for *Pol* to lead *Vpol*.

Table 6. Dividing the main sample into three equal parts after sorting by *income*

Correlation	(1) All 5,872	(2) Smallest third	(3) Middle third	(4) Largest third	Average of (2), (3) and (4)
Correlation between the three main indices					
<i>FHr, Pol</i>	0.897	0.810	0.853	0.919	0.861
<i>FHr, Vpol</i>	0.929	0.834	0.854	0.947	0.878
<i>Pol, Vpol</i>	0.904	0.842	0.882	0.912	0.879
Correlation to <i>Income</i>					
<i>Income, FHr</i>	0.651	0.194	0.169	0.359	0.241
<i>Income, Pol</i>	0.543	0.092	0.161	0.227	0.160
<i>Income, Vpol</i>	0.658	0.182	0.230	0.349	0.254

4.3 *How much do the anchors matter for the correlations?*

It has been proposed that the reason for the high correlations in Tables 4 and 5 is that the anchors are similarly scored in all indices. The upper panel of Table 6 shows that the anchors do not matter much for the correlation between the democracy indices.

The lower panel of the table shows that the anchors matter much more for the correlations between *income* and the democracy indices. It falls substantially when the sample is divided by *income*. To study the Democratic Transition, wide samples or long time series are necessary.

5. The democratic transition in the three main indices

The key result in section 4 was that Factor1 is all democracy indices is strong. It also includes *income* in the Main sample, while income is not included in the OPEC sample. Section 5.1 shows how well income explains the three main indices. Section 5.2 considers the development over time. Sections 5.3 and 5.4 interpret the curves. Section 5.5 estimate the gray zone.

5.1 *Transition curves in unified annual data and country averages*

Figure 3 shows the transition curves in the three samples. The horizontal axis is the same on all three graphs, but only the lower half of the vertical axis is needed on Figure 3c.

The three similar curves on Figure 3a show the democratic transition in the three indices. The functional form of the relation is well determined, as the 95% confidence intervals are narrow; see Net-Appendix. The three curves only differ substantially for incomes below 6.5, where the series have few observations. However, from incomes above 6.5 the three curves all show a perfect transition curve that is flat in the two ends, especially at the top for the HICs (High Income Countries). The remaining five series: *CLr*, *PRr*, *Vlib*, *Vpar*, *Vdel*, and *Vega* all contain the same transition, and in addition it holds in such long data series as are available. Many tests have been made of robustness for the transition relation; see Paldam (2021).

Figure 3b shows the same graph as Figure 3a, but for the 139 country averages. The confidence intervals are a little above 5 pp, so they overlap for most of the range. As the same axes are used, it is easy to see that apart from missing the ends below income 6.7 and above 10.7 the curves look much the same, as they do for the 5-year averages, see Net-Appendix.

Figure 3. Transition curves in the three main indices in C-scale

Figure 3a
Main sample
Unified annual data

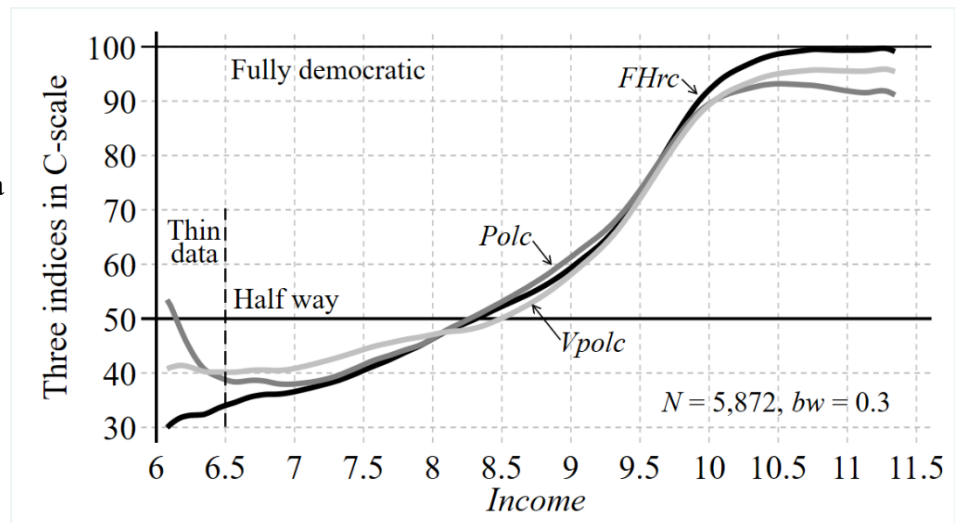


Figure 3b
Main sample
Country averages

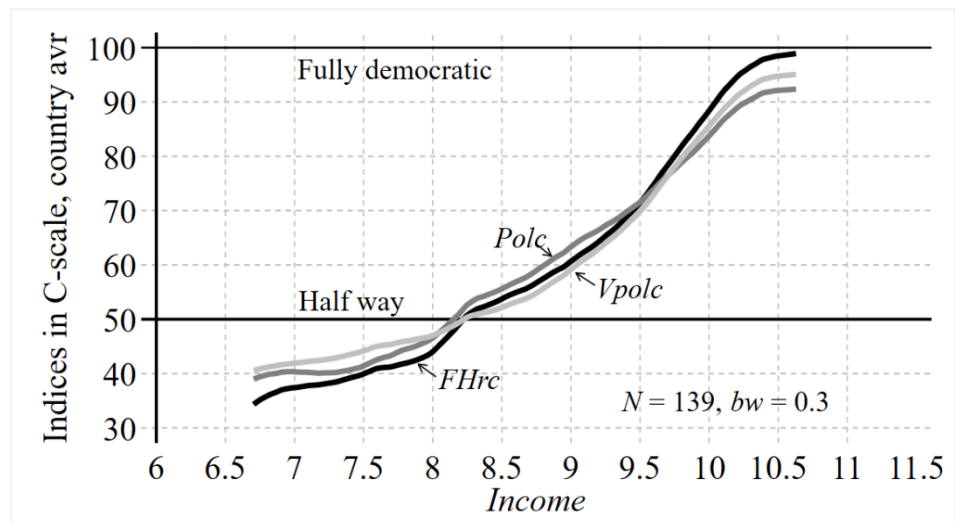
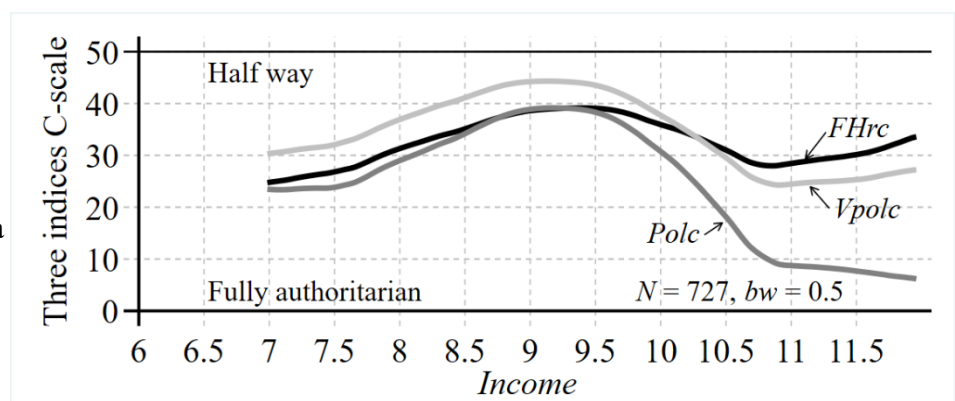


Figure 3c
OPEC sample
Unified annual data



The horizontal income axis is kept the same on all three graphs. The vertical index-axis is the same on Figures 3a and b, but Figure 3c only uses the lower half of the scale. The 95% confidence intervals are omitted; but see Net-Appendix. They are 2-3% on Figure 3a and about 5% on Figures 3b and 3c. Figure 3a has 20 observations below the income of 6.5. It is 0.3% of the observations, so here the data are thin. The Net-Appendix also show all five V-Dem indices and both FH indices.

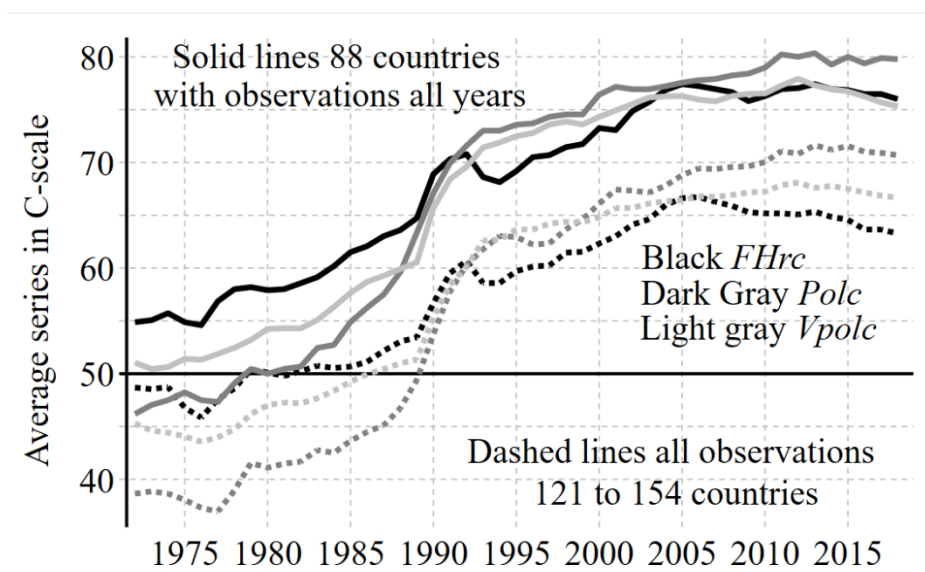
Figure 3c shows the same kernel estimates for the OPEC sample. The sample is smaller, so the bandwidth is larger, and the confidence intervals are wider. The curves are at a lower level throughout, and they do not converge to democracy, but rather to dictatorship. Thus, there is no democratic transition in the OPEC countries. It appears that the transition does start in the income interval from 7 to 9, but then the curves turn down, indicating a turn away from democracy. Finally, the curves become flat at a low level.

Half of the OPEC countries are also in the MENA region (Middle East and North Africa), which consists of Muslim and predominating Arab countries. The curves have a similar path in the MENA countries as in the OPEC countries; see Net-Appendix. Due to the overlap and spatial effects in the Arab group, it is difficult to determine if it is the oil wealth or the Arab/Muslim culture that makes the reaction of the political system to income so different.

5.2 The development over time

Figure 4 shows the development over time. The country sample grows over time from 121 to 154 countries. The largest change is in 1991, where the dissolution of the USSR and Yugoslavia adds 14 new counties. Therefore, each series is shown in two versions. The 88 non-OPEC countries with complete data for all years give the solid lines. The dashed lines are for all observations. The two curves are almost parallel for all three series. The complete series give 10 pp higher curves, as the poorest countries are the ones with most missing observations. On average, the 2x3 indices increase by no less than 25 pp. About half the increase happens around 1990 due to the collapse of Soviet socialism.

Figure 4. The development of the three main indices over time



Since 2005 the curves give an unclear picture. The solid curves show very little either way, while the dashed lines for the expanding sample diverge. There has been much discussion in the media about the ‘crises’ of democracy and the spread of ‘fake news’ through social media, etc. Be it as it may, there is little support in the indices for such drama. What the democracy indices show is that the clear upward trend toward increasing democracy has become much less clear, and maybe it has leveled out. However, the data used ends in 2018, and it is possible that something has happened after that, notably in Russia and Turkey.

5.3 *The theory of the democratic transition: The three pillars and the jumps models*¹⁰

Something that is so strong in the data as the democratic transition needs a general explanation. As all socioeconomic variables have transitions, many explanations can be given. However, we look for basic changes in the **power structure** that generalize to most societies.

Traditional society about 250 years ago in the modern societies of the West and 125 years ago in the modern societies of East Asia all had variants of *the three pillars model*. Power was based on a hereditary king, a feudal nobility, and a national monopoly Church.¹¹ The king was the head of the army, but the nobility, which was about 1% of the population, provided both the top officers and the top clergy. Agriculture produced 50% of GDP. Tenants, who paid rents of about 30% of the production to the feudal owners, did most of the farming. The king and the church were also large landowners. Thus, feudalism was the key institution. Farmers also paid tithe to the Church. The great mass was poor peasants. They were of little consequence in the power structure, and rarely revolted.

The **agricultural transition** reduced the share of agriculture in GDP from about 50% to about 2%. This weakened the wealth and power of the feudal class correspondingly. The weakening allowed land reforms moving property rights to the farmers and abolishing the privileges of the nobility and the tithe. In the same way, the **religious transition** reduced religiosity by about 60%. The land reforms also typically included the land of the Church (and the king). Thus, also the wealth and the power of the Church were much reduced.

This meant that two of the pillars crumbled. Both the nobility and the Church tried to hold on to power in most countries, so the process happened in leaps and bounds. The process is explained by the **jumps model**. It works by triggering events that generate system jumps.

¹⁰ I make no claim to the old three pillars model, but I claim the jumps model; see Paldam (2021), also for evidence supporting the claims of this sub-section.

¹¹ In the new world large-scale farming did start, and some was even done by slaves, so a feudal class was developing in some parts of the two continents. The term Church is used for the institution/organization of a religion.

They happened randomly, and the triggers are quite different. Most are of a political nature. The transition occurs because the transition curve is an attractor for jumps above 15 pp.

As the share of agriculture fell, people moved to towns, where new sectors emerged. After some time, a new middle class emerged. It grew to be the dominating class politically, and it was the main recipient of the large increase in human capital. It wanted mass representation and abolition of the privileges of the nobility, so this all led to democracy and civil rights. Both the *urban transition* and the *transition of education* are strong in the data.

The data contains spatial effects for related countries. This increases the fuzziness of the democratic transition, but it still takes place around a well-defined transition path.

5.4 *Development in the OPEC countries from Figure 3c*¹²

An oil sector is a small, normally heavily fenced, enclave in the economy. It has little impact on the economy except in one respect. Resource rents are easy to tax, so oil export generates an inflow of resource rent into the treasury. OPEC countries produce a lot of oil, so also the flow to the treasury becomes large. In a traditional society, it means that the king becomes rich. It also causes Dutch disease (a real revaluation), so that other exports become uncompetitive in the world market.

With some protection, domestic producers may still produce for the home market, but the country can surely afford to import. To keep power, the king may distribute rents to his friends and clients. Thus, a rentier class may develop, and as everybody knows that they are dependent upon the king, society becomes conservative. In addition, the king can afford a large external and internal security establishment. He can also support the national Church to build a strong alliance. This all gives a different development curve as shown.

It has often been suggested that Islam rather than oil causes the lack of democracy in the OPEC countries as eight of the 16 OPEC countries are Arab/Muslim.¹³ These countries are also a substantial part of the MENA group of countries, i.e., the Middle East and North Africa. The transition curves are similar for the MENA and the OPEC countries, so the two explanations are difficult to sort out. It seems that oil-wealth gives most, but MENA adds something as well. Strong spatial effects make it difficult to sort out the MENA and OPEC effects.

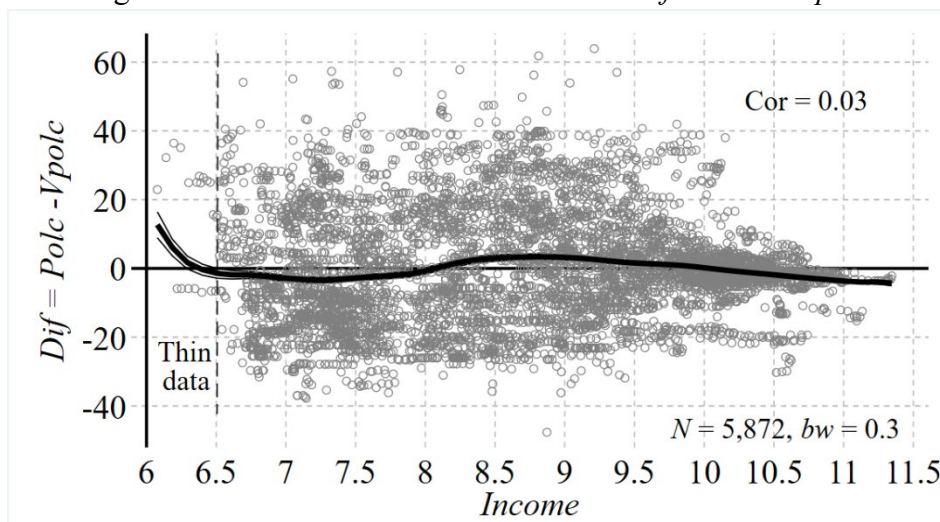
¹² No OPEC country was a high-income country before oil was found, but then income jumped dramatically. If countries are already high-income countries when they find large oil deposits, it is a different story.

¹³ As the reader can imagine, this is a controversial issue, where a literature exists; see Borooah and Paldam (2007), Rowley and Smith (2009) and Potrafke (2012).

5.5 Comparing the *Polc* and *Vpolc* series: Finding the gray zone

Figure 5 shows the average *Dif*-curve is almost horizontal, much as expected from Figure 3. However, Figure 5 also shows how much the 5,872 observations scatter: The average difference between the 5,872 observations for *Polc* and *Vpolc* is zero as it should, but the average numerical difference is 12.4 percentage points.

Figure 5. The non-OPEC observations for $Dif = Polc - Vpolc$



It is obvious that when you compile and code 5,872 observations some errors will occur, but as the indices grow older such errors will be detected and corrected. A large literature exists on coding error. From a previous study, I conclude that at most 0.5 pp can be such error.¹⁴ In addition, the indices have small systematic deviations from linearity so that the kernel curve has some curvature. This also explains a bit of the 12.4 points. However, most of the scatter is no error, but due to the index problem. I conclude that the gray zone is at least 10 pps.

6. The concordance of *Polc* and *Vpolc*, and the largest deviations

This issue is analyzed by two statistics reported for all 155 countries (in the Net-Appendix): (i) $Av\ Dif$, and (ii) the correlation $cor(Polc, Vpolc)$. Section 5.1 looks at *Dif* for six major country groups. Sections 5.2 and 5.3 consider both statistics for all 155 countries.

¹⁴ See Doucouliagos and Paldam (2013). Google Scholar gives more than 3 million hits to ‘coding error’. About 10% are studies of such errors in practice. Having read 25 seemingly relevant papers, I think that one should expect ½% errors. Large errors such as decimal point errors, where 0.235 are coded 2.35, stick out, so they are likely to be detected and corrected, but small random errors, where 0.235 is coded 0.253, may escape detection. However, they matter little for the use of the index.

6.1 Do the indices report country groups differently?

Table 7 show some averages for the six major country groups. The t-ratios say if the difference between the average for the country group and all other countries is statistically significant. It is debatable if it should be based on all observations or the country averages, so both are reported. The *Vpol* index is nicer to Africa and especially to the MENA group, while *Pol* is nicer to Latin America. These differences are highly significant, while the differences between the rating of the Asian, Post Socialist and Western groups of countries are dubious.

Table 7. The average indices for country groups 1972-2018, C-scale

Group	Data	<i>N</i>	<i>Av Income</i>	<i>Av FHrc</i>	<i>Av Polc</i>	<i>Av Vpolc</i>	<i>Av Dif</i>	T-ratio	<i>Av cor</i>	<i>na</i>
Africa	Annual	1,892	7.568	34.96	42.60	45.37	-2.74	-7.0		
	Countries	43					-2.74	-2.1	0.831	1
Asia	Annual	908	8.514	45.68	54.43	52.47	1.96	3.5		
	Countries	20					1.55	0.6	0.778	2
Lat Am	Annual	1,009	8.892	63.28	72.10	66.70	5.40	11.4		
	Countries	22					5.36	2.8	0.724	1
MENA+	Annual	825	9.183	25.49	23.94	32.63	-8.67	-15.7		
	Countries	20					-8.29	-3.1	0.772	4
Post Soc	Annual	884	9.197	51.70	61.57	61.38	0.22	-0.4		
	Countries	27					0.93	-0.3	0.721	5
West	Annual	1,081	10.241	96.17	97.99	98.75	-0.76	-4.7		
	Countries	23					-0.76	-1.1	0.457	14

The *t-test* is two-sided and accepts different variances. It tests whether the observations for *Polc* and *Vpolc* have the same mean. The groups are the standard groups used by international organizations. The correlation is small for the group of West, but in most cases, it is the effect of one small change in the *Polc*. The correlation is *na* (not available) when *Polc* is constant. 14 of the 23 countries in the west group have *Polc* constant at 100.

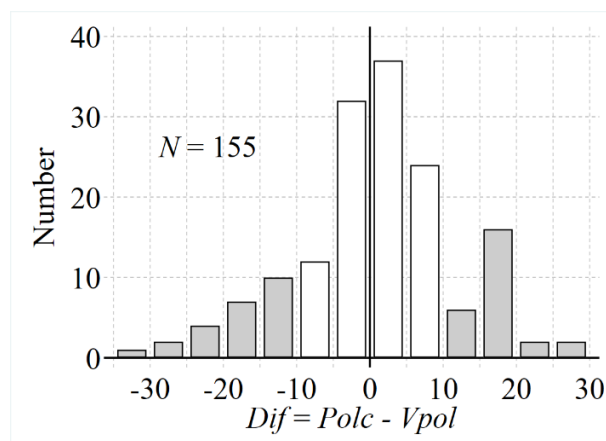
The groups used for Table 7 build on the standard groups used by international organizations: *Africa* is Sub-Saharan Africa. *Asia* includes Mauritius. *Lat Am* is Latin American and the Caribbean. *MENA+* are the Arab countries, Iran, and Turkey, the + means that Afghanistan and Pakistan are included in the group. *Post Soc* is the countries of the old Soviet Block, Yugoslavia, and Albania. *West* includes four overseas western countries. The correlation is small for the group of West, but in most cases, this is the effect of one small change in the *Polc*.

6.2 The concordance between *Polc* and *Vpolc* for the countries

The series for *Vpol* change every year, but most changes are small, and thus highly insignificant. *Pol* is constant most years. The reason that *Vpol* and *Pol* correlate as shown by Figure 6b is thus the large movements. Figure 6a shows that 67.7% $\approx 2/3$ of the countries have *Polc* and *Vpolc* values within one gray zone of 10 pp. Section 6.3 looks at the last 1/3. Figure 6b shows that about 60% of the correlations on Figure 6b are above 0.7. The 17½% *na* (not available) correlations are mostly for western countries where *Polc* gives the score 100 for full democracy and where the *Dif* variable is small. For 20 countries, it is less than five pp.

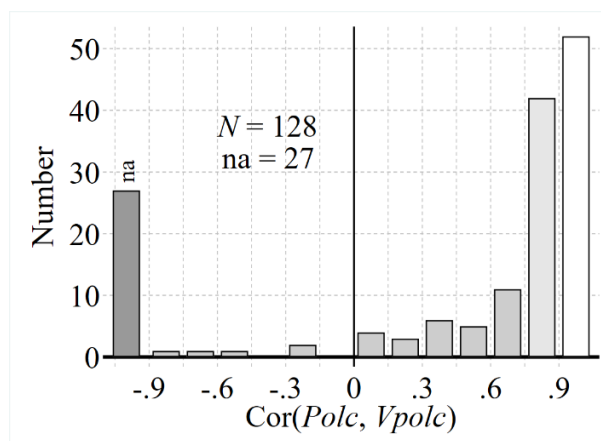
Figure 6. The differences between *Polc* and *Vpolc* for the 155 countries

Figure 6a. Distribution of $Dif = Polc - Vpol$



White bars are for *Dif* below one gray zone

Figure 6b. Distribution of $cor(Pol, Vpol)$

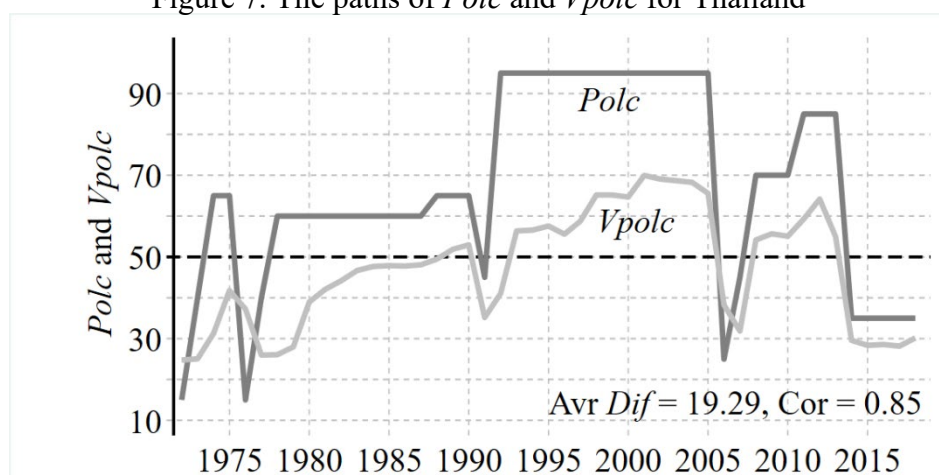


White and light gray bars are for *cor* about 1.

Na, not available, is where *Pol* is constant all years

To get an impression of how the two series track each other, Figure 7 shows the series for Thailand, which has an unusually high regime volatility. It is obvious that the series track each other quite well (the correlation is 0.85). Thai civilian regimes have reasonably good political and civil rights, while the Thai military system suspends such rights. This difference between the two types of regimes is scored much larger by the Polity project.

Figure 7. The paths of *Polc* and *Vpolc* for Thailand



It is also illustrative to compare the two most authoritarian regimes North Korea and Saudi Arabia. *Polc* scores both countries as 0 (though with a small deviation before 1994), while *Vpolc* has flat curves for both countries with North Korea at 22.5 pp and Saudi Arabia at

15.2 pp. The explanation may be that North Korea has a fine constitution that formally grants all political and civil rights, but it has no reality. Saudi Arabia has the Koran as its constitution. Both regimes are repressive, but it appears that North Korea takes the price for repression.

6.3 Some cases of poor concordance

Malaysia has the largest value of the average *Dif*; it is 29.1 pp as shown on Figure 8. The indices have a reasonable correlation but a different level. It is surprising that *Vpol* places Malaysia below the mid-level of 50 pp. The country has two ethnic/religious groups. The political elite is from the Malay majority, and the economic elite is from the substantial Chinese group. The groups have a complex relation requiring many compromises.

Figure 8. The paths of *Polc* and *Vpolc* for Malaysia

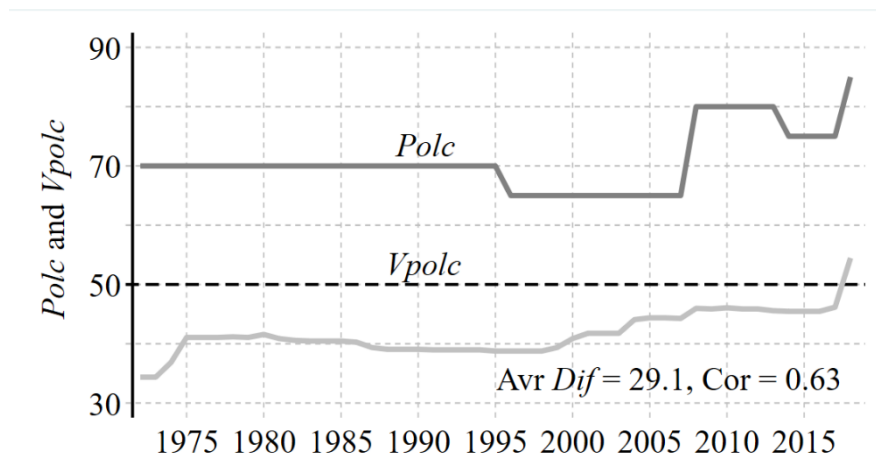


Figure 9. The paths of *Polc* and *Vpolc* for Colombia

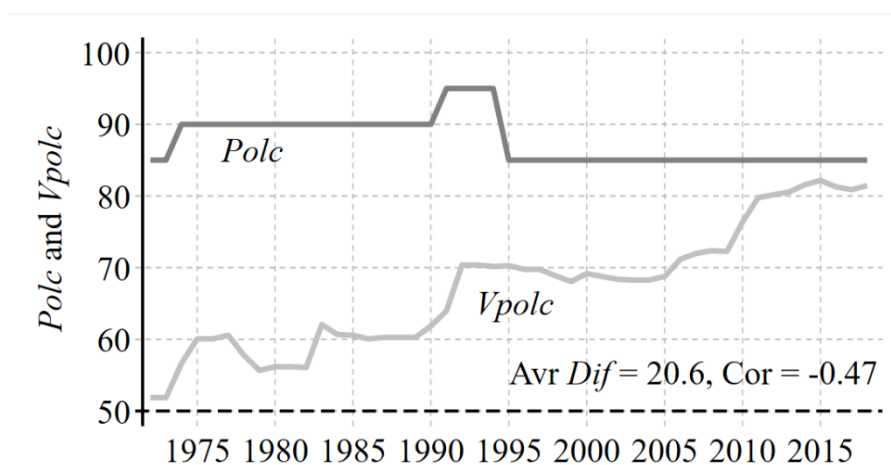
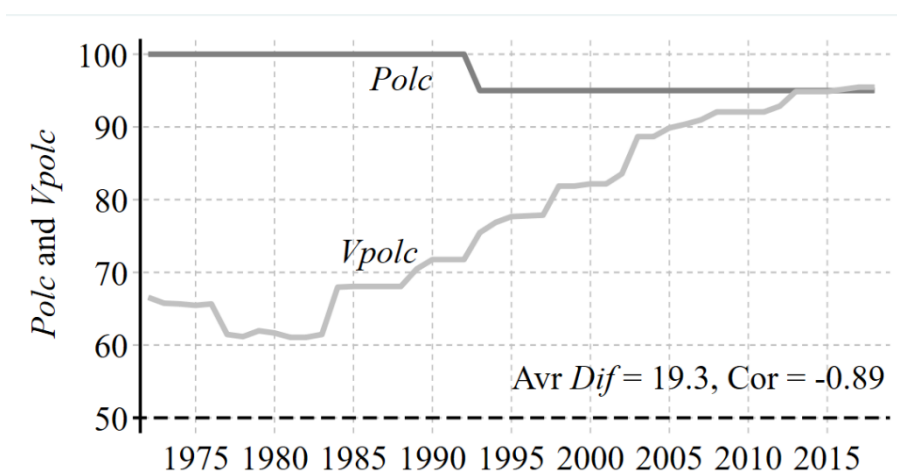


Figure 10. The paths of $Polc$ and $Vpolc$ for Jamaica



For Colombia and Jamaica, both Dif and cor are outliers. It is hard to understand that two groups of solid researchers can score two countries so differently.

$Polc$ and $Vpolc$ also differ for: (i) El Salvador, where for a period the difference was almost 50 pp. (ii) Iran, where the democratic system has a theocratic super-system. (iii) Israel, where it is hard to determine which area should be covered. (iv) South Africa, where the deviation was quite big in the apartheid period.

7. Conclusions

The paper has reached three conclusions:

1. All eight indices have the same grand pattern, dominated by the democratic transition. The transition is an underlying long-run process, which is rather fuzzy in the short run in each country, but strong and robust in large datasets, both across countries and over time. The OPEC/MENA countries constitute an important exception. This means that studies of large data samples using a democracy index are robust to the choice of the index.

2. The strong transition means that long-run causality is mainly from development to the political system. Recall the discussion of the Jumps Model where stochastic chocks to the system trigger jumps in the direction of the transition path. Thus, we are not dealing with a clean short-run causal link, but with an underlying long-run causal link.

3. The indices are only measured up to a gray zone of at least 10%. It is due to the index problem and not to mistakes. Thus, when two democracy observations differ by less than 10% the difference should be considered insignificant. Consequently, researchers who work with small country groups should be sure to check results reached with one index by a replication

with one or two other indices.

Two criteria of concordance are calculated for the two main indices *Polc* (Polity) and *Vpolc* (Polyarchy): (i) the indices differ less than the gray zone of 10 pp in 2/3 of the countries. (ii) The correlation between the two indices is below 0.7 in about one third of the countries. Though the two groups of deviators have some overlap, it still means that about 50% of the countries deviate by at least one of the two criteria.

References:

- Arrow, K.J., 1963. *Social Choice and Individual Values* (2nd ed.). Yale UP, New Haven
- Barro, R.J., Sala-i-Martin, X., 1995, 2004. *Economic Growth*. MIT press, Cambridge MA
- Bjørnskov, C., Rode, M., 2020. Regime types and regime change: A new dataset on democracy, coups, and political institutions. *The Review of International Organizations* 15, 531-51
- Boese, V., 2019. How (not) to measure Democracy. *International Area Studies Review* 22, 95-127
- Borooah, V.K., Paldam, M., 2007. Why is the World Short of Democracy? A Cross-Country Analysis of Barriers to Representative Government. *European Journal of Political Economy* 23, 582-604
- Christoffersen, H., Beyeler, M., Eichenberger, R., Nannestad, P., Paldam, M., 2014. *The good society. Denmark and Switzerland*. Springer, Berlin
- Coppedge, M. et al.¹⁵, 2020. *V-Dem Codebook v10. Varieties of Democracy (V-Dem Project)*. University of Gothenburg V-Dem Institute
- Doucouliagos, H., Paldam, M., 2013. The robust result in meta-analysis of aid effectiveness: A response to Mekasha and Tarp. *Journal of Development Studies* 49 (4), 584-87
- Eichhorn, W., Voeller, J., 1976. *Theory of the Price Index*. Springer, Berlin
- Freedom House democracy index home page: <https://freedomhouse.org/countries/freedom-world/scores>
- Gründler, K., Krieger, T., 2021. Using Machine Learning for measuring democracy: A practitioner's guide and a new updated dataset for 186 countries from 1919 to 2019. *European Journal of Political Economy* 70, 102047
- Gründler, K., Krieger, T., 2022. Should we care (more) about aggregation? *European Economic Review* 142, 104010
- Gundlach, E., 2021. A note on democratic transitions and humps. Working paper
- Jones, C.I., Vollrath, D., 1998, 2013. *Introduction to Economic Growth*. Norton, New York
- Maddison project, home page <https://www.rug.nl/ggdc/historicaldevelopment/maddison>
- Marshall, M.G., Gurr, T.R., Jaggers, K., 2018. *PolityTM IV project. Political Regime Characteristics and Transition. Dataset Users' Manual*. Center for Systemic Peace
- Munck, G.L., Verkuilen, J., 2002. Conceptualizing and Measuring Democracy: Evaluating Alternative Indices. *Comparative Political Studies* 35, 5-34
- Paldam, M., 2021. *The Grand Pattern of Development and the Transition of Institutions*. Cambridge UP, Cambridge UK and New York (contains a detailed reference list of seven pages)
- Paldam, M., 2022. Net-Appendix to the present paper: <http://>
- Polity project, home page <https://www.systemicpeace.org/polityproject.html>
- Potrafke, N., 2012. Islam and democracy. *Public Choice* 151, 185-92
- Rowley, C.K., Smith, N., 2009. Islam's democracy paradox: Muslims claim to like democracy, so why do they have so little? *Public Choice* 139, 273-299
- V-Dem home page: <https://www.v-dem.net/en/login>

¹⁵ The co-authors/research group: J. Gerring, C.H. Knutsen, S. Lindberg, J. Teorell, D. Altman, M. Bernhard, M.S. Fish, A. Glyn, A. Hicken, A. Lührmann, K.L. Marquardt, K. McMann, P. Paxton, D. Pemstein, B. Seim, R. Sigman, S.-E. Skaaning, J. Straton, A. Cornell, L. Gastaldi, H. Gjerløv, V. Mechkova, J. von Römer, A. Sundström, E. Tzeglov, L. Uberti, Y-t. Wang, T. Wig, D. Ziblatt.