

*Online Appendix*  
Democracy for All:  
Conceptualizing and Measuring  
Egalitarian Democracy

Rachel Sigman and Staffan I. Lindberg

January 2018

## A Index Construction

The information in this section supplements the information found in Section 3. Table 1 provides descriptive statistics for the 10 V-Dem indicators used to construct the egalitarian indices. Table 2 provides summary statistics for the indices resulting from the Bayesian Factor Analyses.

**Table 1:** Descriptive Statistics for Constituent V-Dem Indicators

Indicator	Mean	SD	N
<b>Equal Protection</b>			
Social class equality in civil liberties	0.069	1.504	16181
Social group equality in civil liberties	0.034	1.472	16620
Weaker civil liberties (% of population)	42.69	20.642	15408
<b>Equal Distribution</b>			
Particularistic or public goods	0.015	1.483	16620
Universalistic welfare policies	-0.174	1.471	16620
Educational equality	-0.345	1.631	16620
Health equality	-0.236	1.597	16,620
<b>Equal Access</b>			
Power distributed by socioeconomic position	-0.405	1.448	16620
Power distributed by social group	-0.207	1.516	16620
Power distributed by gender	-0.793	1.322	16620

**Table 2:** Egalitarian Indices Descriptive Statistics

Variable	Mean	SD	Min	Max	N
Equal Protection	0.500	0.289	0.021	0.991	14969
Equal Distribution	0.488	0.297	0.038	0.991	16620
Equal Access	0.495	0.292	0.023	0.997	16620
Egalitarian Component	0.483	0.266	0.056	0.990	14969
Electoral Democracy index	0.321	0.281	0.008	0.958	16259
Egalitarian Democracy	0.248	0.230	0.017	0.919	14643

### Egalitarian Democracy Index Aggregation Formula

The underlying rationale for the aggregation formula presented below is that equal weighting of the additive terms and the multiplicative term respects both the Sartorian ‘necessary condition’ logic and a ‘family resemblance’ logic (?). To some extent, the degree of egalitarianism still matters even when there is no electoral democracy, and electoral democracy still matters even when there is no egalitarianism. But there are strict boundaries to how much each dimension matters in isolation (for further details including the detailed rationale for weights, see ? and ?. Thus, if a country scores 0 (the minimum) on Electoral Democracy and 1 (the maximum) on the Egalitarian

Component index, its score on the Egalitarian Democracy index would be .25, thereby recognizing that the country possesses some of the attributes essential for egalitarian democracy but could not be considered a democracy.<sup>1</sup>

$$v2x\_egaldem = .25*v2x\_polyarchy^{1.6} + .25*v2x\_EgalComp + .5*v2x\_polyarchy^{1.6}*v2x\_EgalComp$$

## B Content Validity

### Index and Indicator Correlation

A correlation matrix of the five indices (Table 3) aids us in assessing the coherence of the indices. Overall, the weakest correlations are found between the Egalitarian Democracy index and the three subcomponent indices suggesting, again, that electoral democracy alone does not necessarily capture attributes related to egalitarianism. The Egalitarian Component index correlates both strongly and similarly with all three subcomponents, and the three subcomponents correlate with each other at a level of .75-.80. The relationship between the subcomponents is especially encouraging, since these associations point to the presence of mutually-reinforcing dynamics between these three important egalitarian concepts.

**Table 3:** Correlation Matrix of Egalitarian Indices\*

Index	Egal. Dem.	Elec. Dem.	Egal. Comp.	Eq. Protec.	Eq. Dist.	Eq. Acc.
Egal. Dem.	1.000					
Elec. Democ.	0.954	1.000				
Egal. Comp.	0.857	0.747	1.000			
Eq. Protec.	0.776	0.661	0.920	1.000		
Equal Dist.	0.786	0.668	0.928	0.784	1.000	
Equal Acc.	0.812	0.738	0.917	0.760	0.788	1.000

\*All correlation coefficients statistically significant at 99% confidence level

In Section 4.1 we presented factor loadings and uniqueness scores from the results of the Bayesian Factor Analysis. As a supplement those statistics, we present, in Table 4, correlation coefficients of each constituent indicator with each of the five indices. In similar fashion to the loadings and uniqueness scores, the correlations coefficients show that the indices tend to correlate most strongly with measures related to socioeconomic class and the equal distribution of health and educational resources for the Equal Distribution index. The lowest correlations, as expected, are with the indicator measuring the percent of the population for whom civil liberties are not equally protected. Given that this variable is rated on a scale of 0 to 100 (rather than the ordinal ratings of the other indicators) it is not surprising to find lower levels of correlation as there is likely to be more variance in ratings.

<sup>1</sup>One could argue that electoral democracy should be viewed as a "hard" necessary condition of egalitarian democracy and we should, therefore, not include the additive component of the aggregation. Since we provide the constituent indices, it will be easy for anyone to do this should they disagree with our aggregation.

**Table 4:** Correlations of Egalitarian Index and Constituent Indicators

Indicator	Eq. Protec	Eq. Dist.	Eq. Acc	Egal Comp.	Egal Dem.
Social Class Equality in Respect for Civil Liberties	0.975	0.766	0.768	0.896	0.755
Social Group Equality in Respect for Civil Liberties	0.825	0.672	0.667	0.773	0.687
Equal Civil Liberties % of Population	-0.609	-0.593	-0.479	-0.604	-0.611
Particularistic or Public Goods	0.677	0.719	0.702	0.761	0.666
Universal or Means-Tested Welfare	0.601	0.703	0.654	0.705	0.590
Educational Equality	0.728	0.948	0.752	0.875	0.754
Health Equality	0.762	0.969	0.752	0.896	0.790
Power Distributed by Socioeconomic Class	0.710	0.741	0.918	0.849	0.709
Power Distributed by Social Class	0.678	0.681	0.866	0.798	0.753
Power Distributed by Gender	0.615	0.683	0.856	0.775	0.740

Number of observations range between 14969 and 16620

All correlation coefficients statistically significant at 99% confidence level

## Pooled Variance Decomposition

A pooled model of variance decomposition (Table 5) enables us to analyze the coder-level variance components for an entire index. In this model the grand mean is calculated not from an individual indicator but from ratings across all indicators included in each index. Again, this model employs fixed effects for country- and year-variation and, this time, crossed random effects for coder- and indicator-levels. The crossed-effects specification reflects the non-nested structure of the relationship between coders and indicators in the V-Dem data: coders may code multiple, but not all, indicators.

The estimates of these models reveal differences in coder-level variance across the indices. Coder-level variance is highest for Equal Protection (0.043) and lowest for Equal Access (0.023). These can be interpreted to indicate the relative clarity and coherence of the principle of Equal Access, and the fact that the indicators are more closely related, when compared to Equal Distribution and Equal Protection. Coder-level variance for the Egalitarian Component index is on the lower end of this spectrum, at 0.029. Yet, in all cases, the variance associated with coder disagreement is considerably lower than the indicator-level variance component. We interpret this to mean that conceptual ambiguity or confusion among raters is not seriously undermining the content validity of our measures.

**Table 5:** Coder Disagreement Pooled by Index

Indicator	Coder Effects	SE	Ind.	SE	Coders	Obs.
Equal Protection	0.043*	0.002	0.067*	0.000	1,192	208,569
Equal Distribution	0.038*	0.001	0.059*	0.000	1,654	331,168
Equal Access	0.023*	0.001	0.051*	0.000	1,160	249,357
Egalitarian Comp.	0.029*	0.001	.069*	0.000	1,936	789,094

\*p<.01

## Coder Characteristics and Bias in Ratings

A potential source of bias in ratings could come from coders with particular background characteristics, such their gender, ideology, or level of education. Table 6 shows the effects of coder characteristics on index variance as predicted by a multi-level model that, instead of looking at overall coder variance, breaks down into coder attributes. Women and those employed by the government are indeed more likely to assign lower scores (less egalitarian) on the various indicators, though the size of the effect varies across these different groups and across indicators, and is generally small. This trend is not too surprising, however, given that these groups may be more likely to have experienced the effects of unequal protections or distributions. Experts who are either students or describe themselves as believing in the free market are slightly more likely to assign higher scores (more egalitarian). There appears to be no statistically significant effect of whether or not the expert holds a PhD, and there is a statistically significant but substantively questionable effect for coders who reside in the country.

Overall, the small size of the variance generated by each type of coder is reassuring.<sup>2</sup>

**Table 6:** Effects of Coder Characteristics

	Eq. Protec.	Eq. Dist	EqAcc	Eg. Comp
PhD	-0.011 (0.017)	0.006 (0.006)	-0.003 (0.006)	-0.002 (0.006)
Gov	-0.007 (0.014)	-0.060** (0.025)	-0.067*** (0.010)	-0.050*** (0.013)
Female	-0.002 (0.008)	-0.025** (0.010)	-0.003 (0.006)	-0.011* (0.006)
Freemarket	0.009* (0.005)	0.010*** (0.003)	0.015*** (0.003)	0.011*** (0.002)
Resident	-0.000** (0.000)	0.000*** (0.000)	-0.000 (0.000)	0.000*** (0.000)
Time	0.001** (0.000)	0.000 (0.000)	0.001** (0.000)	0.001*** (0.000)
Student	0.002 (0.003)	0.033** (0.013)	0.013*** (0.004)	0.018** (0.006)
<b>Variance Components</b>				
Indicator	0.002*** (0.001)	0.000*** (0.000)	0.003*** (0.002)	0.003*** (0.002)
Country	0.044*** (0.002)	0.047** * (0.008)	0.032*** (0.007)	(0.041)*** (0.004)
Obs	175,792	287,098	217,443	680,333
Coders	1,192	1,654	1,160	1,936

\* p<.1 \*\*p<.05 \*\*\*<.01

## Inter-Rater Agreement

As stated in section 4.2, we provide more traditional measures of inter-rater agreement (IRA), including Cohen's kappa (weighted and unweighted versions) and Krippendorff's Alpha. While Cohen (1960) suggested that any kappa statistic above .8 represented "substantial agreement" there is little precedent against which to assess the agreement levels in the V-Dem data. We thus focus principally on comparisons of levels of rater agreement amongst the indicators in the same way we do in the variance components analysis above. Unlike the variance components analysis, however, it is not, to our knowledge, possible to pool these analyses for each index taking multiple levels into account.<sup>3</sup>

Table 7 shows common measures of inter-rater agreement for each of the nine

<sup>2</sup>While the existence of effects alone may be cause for mild concern in terms of consistency in coder ratings, they also suggest that V-Dem's approach of recruiting diverse groups of coders is potentially very important for understanding the issues involved in the production of expert survey data.

<sup>3</sup>Additionally, there is considerable disagreement about the relative strengths and weaknesses of different IRA statistics. For an overview, see Gwet (2014).

ordinal variables used to construct the indices.<sup>4</sup> The first column is an unweighted Kappa statistic that does not assume constant identity of raters across all ratings. The Kappa statistic calculates a proportion of observed agreement relative to how much agreement would be expected to be present by chance alone, based on the number raters, ratings and number of categories. For each indicator, it takes the average across each of the four ordinal categories. The second and third columns display weighted versions of the Kappa statistic where differences in proximate ratings (i.e. 1 and 2 or 3 and 4) are weighted to reflect more agreement than would less proximate ratings (i.e. 1 and 4). The second column assigns weights  $1 - |i - j|/(k - 1)$ , where  $i$  is country-year,  $j$  is the coder and  $k$  is the maximum number of possible ratings. The third column is an alternative weighting specification  $1 - (|i - j|/(k - 1))^2$  which effectively weights smaller differences in ratings more than larger differences in ratings. The final column shows Krippendorff's alpha (1980). The main difference between Krippendorff's alpha and the Kappa statistics is that Krippendorff's alpha measures disagreement, whereas Kappa statistics measure agreement.

**Table 7:** Measures of Inter-Rater Agreement by Indicator

	$\kappa$	$\kappa(w)$	$\kappa(w^2)$	Krip. $\alpha$
Equal Civ. Lib. - Socioeconomic Class	0.140	0.129	0.190	0.428
Equal Civ. Lib. - Social Group	0.142	0.137	0.200	0.386
Particularistic or Public Goods	0.125	0.128	0.179	0.327
Universal or Means-Tested Welfare	0.138	0.162	0.206	0.314
Educational Equality	0.225	0.217	0.298	0.541
Health Equality	0.228	0.232	0.326	0.545
Power Distributed by Socioeconomic Class	0.161	0.190	0.260	0.418
Power Distributed by Social Class	0.219	0.186	0.252	0.450
Power Distributed by Gender	0.192	0.266	0.372	0.472

All correlation coefficients statistically significant at 99% confidence level

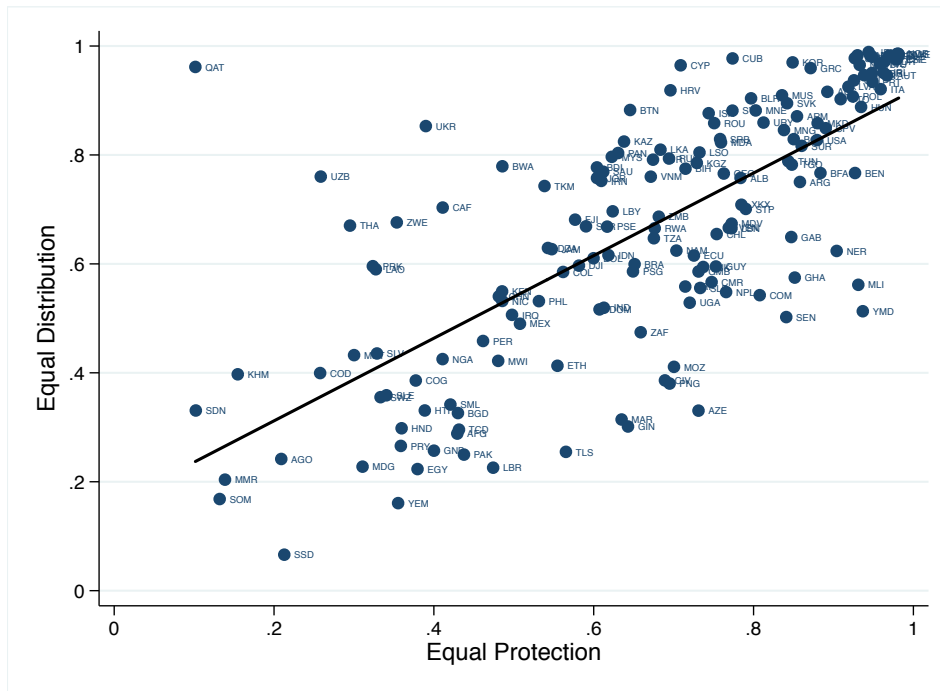
Overall, the levels of IRA are not considerably high which could be cause for concern in terms of assessing the extent to which raters have common understandings of the concepts being measured. That said, it is difficult to know what levels of IRA should be expected with this type of data. Moreover we do see fluctuation in IRA levels that are similar to that observed in the variance components analysis in section 4.2. In particular, the indicators used in the Equal Protection index have the lowest levels of IRA while measures of health and education equality and power distribution - those that are most observable - demonstrate higher levels of IRA.

## C Face Validity

Figure 1 plots the Equal Distribution and Equal Protection subcomponents using average (mean) scores over the period of 1990-2012. As expected, the scatter plot shows a strong positive correlation between these two dimensions. In general, countries that

<sup>4</sup>As noted above, the data that were actually used in the construction of indices were point estimates from the V-Dem measurement model output, but the data were originally collected in ordinal format.

**Figure 1:** Scatter Plot: Equal Protection vs. Equal Distribution 1990-2012



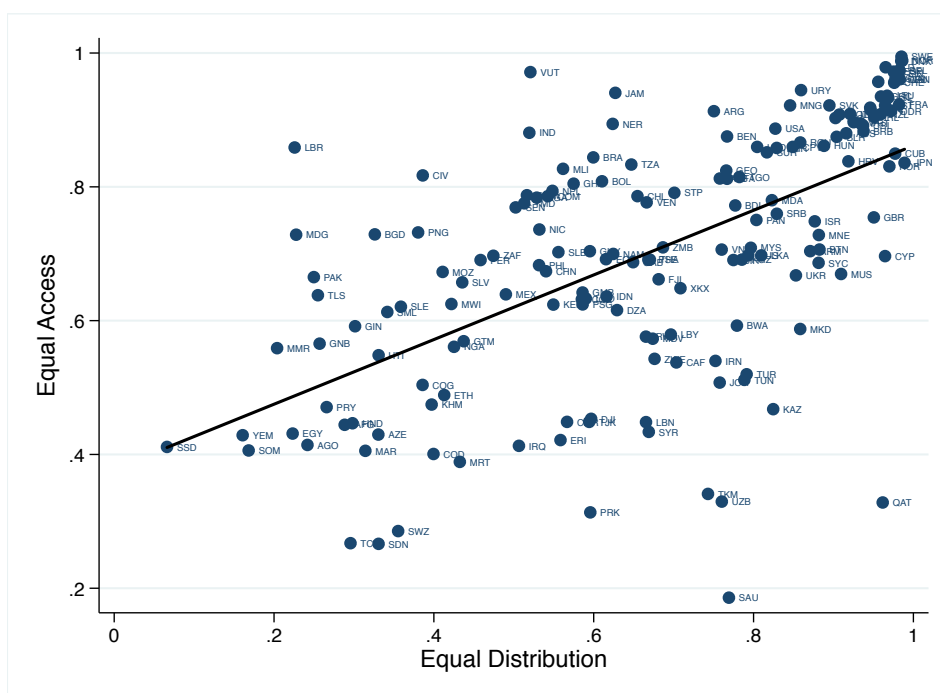
we would expect to score high on both dimensions are found in the upper right-hand corner, such as the Nordic countries and post-communist democracies such as Poland, the Czech Republic and others. Communist countries like Cuba, China and North Korea demonstrate relatively high Equal Distribution scores for their levels of Equal Protection, while those countries with relatively liberal regimes but weaker distributions of resources – Mali and Yemen for example – appear further below the best-fit line.

Figure 2 plots Equal Distribution and Equal Access averaged over the same period. The relationships between these two variables appears slightly less strong with, once again, a cluster of highly egalitarian countries in the upper righthand corner. Well above the best fit-line are a number of African countries where relatively equal access to power does not coincide with more equal distributions of resources. India also falls in this category. Well below the best-fit line we find a mix of countries, mainly in Central Asia and the Middle East, where equality in the distribution of resources has outpaced equal access to power. For these countries, the relationship appears to be capturing a dynamic whereby a small group of powerholders distributes oil rents or other resources in order to maintain unequal access to power.

Time trends in the subcomponent indices in the United States are presented in Figure 3. Equal Protection sees its steepest rise around the time of the 1964 Civil Rights Act, though the measure does not seem to capture the fact that *de facto* protections were slower to take effect in Southern states. Equal Distribution rises following the passage of the New Deal in the 1930s and again in the 1960s, but levels off as retrenchment of these major social programs becomes more prevalent in the 1970s and 1980s. The Equal Access subcomponent sees major shifts in access to power in the 1920s and 1930s following the enfranchisement of women and again in the late 1960s. In recent decades (up to 2012), as expected, Equal Distribution remains considerably lower than Equal



**Figure 2: Scatter Plot: Equal Distribution vs Equal Access 1990-2012**



Protection and Equal Access, explaining why scores for the U.S. on the Egalitarian Democracy index are lower than many other advanced industrialized countries.<sup>5</sup>

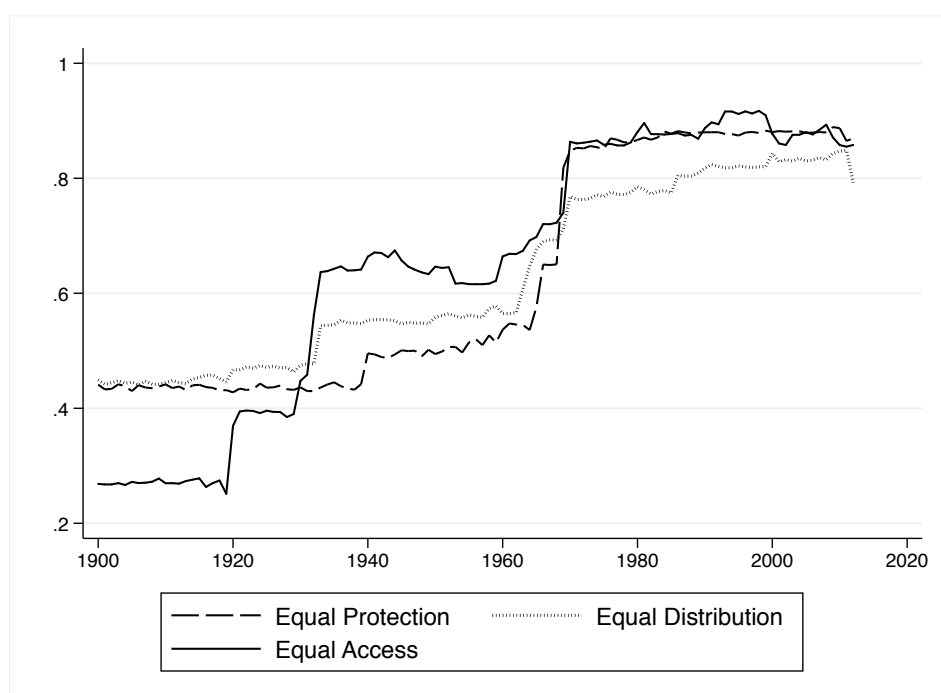
## D Convergent/Discriminant Validity

Table 8 shows correlations of the Egalitarian indices with a selection of V-Dem's other indices. With some exceptions, these correlations tend to fall in the .6-.8 range and vary in expected ways.<sup>6</sup> Negative and weaker correlations with corruption are comforting given the large conceptual gulf between equality and the behavior of public officials. The relatively weaker associations with the suffrage index are also supportive of our measures given that the mere existence of formal voting rights has less in common with egalitarianism than, say, participation or deliberation. Although the Participatory Component index correlates more strongly with the Egalitarian Democracy index, its weaker correlation with the other three indices is, at first glance, somewhat concerning given the centrality of participation in our conceptualization of egalitarian democracy. A closer examination of V-Dem's Participatory Component index, however, sheds light on why we do not see a stronger correlation. Among the indicators included in the participatory index are measures of direct popular vote (v2xdd\_dd), elected local

<sup>5</sup>It's important to note that, by design, the equal distribution subcomponent does not track directly with measures of income inequality such as GINI coefficients. The GINI index of income inequality, according to OECD data, increased from approximately .31 in 1979 to .39 in 2012 for the United States while our distribution measure shows a slight increase up until 2011 and then a decrease in 2012 when V-Dem's measures of universal welfare and health equality begin to decrease, perhaps due to the prevalence of states restrictions in these areas.

<sup>6</sup>We expect stronger correlations with the Egalitarian Democracy index due to the fact that it includes a number of indicators that are also included in other indices.

**Figure 3: Equal Protection, Access and Distribution in the United States**



government (v2xel\_locelec), and elected regional government power(v2xel\_regelec), all of which measure only formal opportunities for voting rather than concepts related to the extent to which various groups in society can actually take advantage of these opportunities.

Although, to our knowledge, no other measures of egalitarian democracy exist, we can examine how our indices compare to measures of concepts related to the principles of egalitarian democracy as we have outlined them above. We highlight several of these comparisons in Table 9.

First, looking at Freedom House's measures of Civil Liberties and Political Rights, we see that there is a clear distinction between these measures and the egalitarian

**Table 8: Correlations with Other V-Dem Indices**

Index	Eq. Prot.	Eq. Dist.	Eq. Acc.	Egal	Egal Dem
Polyarchy	0.661	0.668	0.738	0.747	0.954
Participatory	0.603	0.611	0.694	0.692	0.824
Liberal	0.615	0.580	0.604	0.648	0.807
Deliberative	0.639	0.653	0.717	0.724	0.838
Corruption	-0.361	-0.372	-0.307	-0.353	-0.430
Civil Society Participation	0.607	0.590	0.701	0.685	0.826
Free Expression	0.537	0.511	0.611	0.599	0.802
Gender Empowerment	0.721	0.760	0.831	0.824	0.841
Suffrage	0.569	0.613	0.699	0.678	0.569

All correlation coefficients statistically significant at 99% confidence level

**Table 9: Correlations with Other Measures**

	Eg. Prot.	Eg. Dist	Eq. Acc	Egal	Egal Dem
Civil Liberties (FH)	-0.539	-0.484	-0.603	-0.609	-0.848
	5596	6277	6277	5596	5584
Political Rights (FH)	-0.510	-0.463	-0.611	-0.593	-0.851
	5596	6277	6277	5596	5584
Worker Rights (Ciri)	0.422	0.335	0.443	0.442	0.550
	3986	4459	4459	3986	3985
Income Inequality (Gini)	-0.498	-0.632	-0.557	-0.604	-0.651
	9766	10950	10950	9766	9590
Average Education (Barro & Lee)	0.610	0.767	0.657	0.728	0.751
	11929	13294	13294	11929	11682
Infant Mortality	-0.454	-0.642	-0.519	-0.577	-0.611
	9972	11221	11221	9972	9794
Non-Resource Tax (ICTD)	0.560	0.699	0.567	0.672	0.670
	3620	4111	4111	3620	3618
GDP Per Capita	0.4418	0.574	0.448	0.520	0.665
	9383	10444	10444	9383	9197

Number of observations in parentheses

All correlation coefficients statistically significant at 99% confidence level

indices, which is not surprising given Freedom House's distinctively liberal disposition.<sup>7</sup> Correlations with CIRI's measure of worker rights are among the lowest of the measures examined, suggesting, perhaps, that the protection of worker rights represents a weakness in our own measure or that this concept is not, empirically speaking, as closely related to egalitarian foundations of democracy. In terms of the income inequality Gini, Equal Distribution correlates, as expected, at higher levels than the other indices. This is also true for measures of average years of schooling (Barro and Lee, 2013), infant mortality and non-resource taxes as a percentage of GDP.<sup>8</sup> To the extent that wealthier countries may benefit from more easily providing basic necessities and having the resources necessary to enforce rights and freedoms, we also include a measure of GDP per capita which, perhaps somewhat surprisingly, correlates most strongly with the broader Egalitarian Democracy index and, second, with the Equal Distribution index.

<sup>7</sup>Not surprisingly, there is a stronger correlation with the Egalitarian Democracy index which includes indicators more closely related to the same concepts measured by Freedom House variables.

<sup>8</sup>The tax measure comes from the International Center for Development and Taxation .

## References

- Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics* 104, 184–198.
- Cohen, J. (1960). Kappa: Coefficient of concordance. *Educ. Psych. Measurement* 20, 37.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.
- Krippendorff, K. (1980). Validity in content analysis.