Marketplace of indicators: inconsistencies between country trends of measures of the rule of law

Marta Kołczyńska & Paul-Christian Bürkner

View supplementary material

Published online: 01 Nov 2021.

Submit your article to this journal

View related articles

View Crossmark data

NOTE

Marketplace of indicators: inconsistencies between country trends of measures of the rule of law

Marta Kołczyńska [ID][a] and Paul-Christian Bürkner [ID][b]

[a]Department of Research on Social and Institutional Transformations, Institute of Political Studies of the Polish Academy of Sciences, Warszawa, Poland; [b]Cluster of Excellence SimTech, University of Stuttgart, Stuttgart, Germany

ABSTRACT
Social scientists can choose from among multiple quality of governance indicators which use different conceptualizations of governance and its components, rely on different data sources, and employ different aggregation and scaling techniques. Despite all differences, these indicators are commonly found to be strongly correlated, which makes the choice of indicator for a given analysis seem inconsequential. We focus on rule of law indicators to demonstrate that correlations among them are indeed high when comparing across countries or using pooled country-year data sets, but are surprisingly low – sometimes even negative – within countries. Given the increased interest of researchers in longitudinal analyses with country time series, low agreement between country time trends in the rule of law is concerning. We illustrate the problem with an analysis of the effect of rule of law on popular support for democracy, which leads to opposite conclusions depending on which measure of rule of law is used.

Introduction

The debate about the validity and comparability of governance indicators dates back at least several decades to the critique of existing measures of democracy by Bollen (1980), who also noted that conclusions about the association between inequality and democracy depended on the choice of democracy indicator. Since then, others have examined existing indicators of governance generally finding a large degree of agreement between indicators but also pointing to deviations that may affect model results (Boese 2019; Högström 2012; Malito 2014; Skaaning 2009; Møller and Skaaning 2011).

In recent years, new approaches to creating cross-national indicators of governance yielded new data sets providing measures of governance, often with global coverage and dating back decades if not centuries. These new indicators created by the Varieties of Democracy project, the Democracy Barometer and the Global State of Democracy

project, have joined World Bank's Worldwide Governance Indicators (WGI) and the older Freedom House index and Polity to extend the infrastructure for comparative social science research. The marketplace of indicators has thus grown bigger offering multiple approaches to measuring at least nominally the same concepts. In the perfect situation, as in the idealized marketplace of ideas, the competition would lead to the domination of the superior indicators.

The new projects use elaborate conceptualizations, rich and diverse data sources, and complex modelling techniques to arrive at country-year estimates of different dimensions of governance. This embarrassment of riches leaves researchers with a choice of the most suitable measure for their analysis, while oftentimes the social science theory that provides the framework for the empirical analysis is not precise enough to guide this decision. The choice may not seem consequential if the differences between indicators of the same concept from different projects are small. Whether differences are small, however, depends on the type of analysis one wants to perform. For much of comparative social science research that focuses on differences between countries, high cross-country correlations between indicators are crucial. Validation of new indicators of governance against already existing measures also typically involves cross-national correlations (cf. Lührmann, Marquardt, and Mechkova 2020). When the analysis consists in modelling changes over time, it is necessary to establish whether different governance indicators exhibit the same or similar trends.

In this note, we contribute to the discussion of the comparability of governance indicators by drawing attention to the correspondence between country trajectories. Focusing on the rule of law, we demonstrate that most popular indicators are strongly correlated when looking at pooled country-year data as well as when correlating values for a single time point from different countries. At the same time, within-country (over time) correlations, i.e. correlations between the trajectories of two rule of law indicators from the same country over the same period of time, are typically much smaller and sometimes negative. This means that, in some countries, one rule of law indicator points to an improvement while another rule of law indicator – in the same country over the same period of time – indicates a decline. Similar patterns had been earlier observed by Cope, Crabtree, and Fariss (2019) with regard to indicators of state repression from the Human Rights Data Project and Varieties of Democracy projects, and Standaert (2015) in the case of corruption indicators from the WGI and Corruption Perception Index.

In addition to documenting discrepancies among indicators referring to the rule of law from four datasets commonly used by social scientists, we show that these discrepancies can have serious consequences for statistical analyses, especially those focused on explaining change. As an illustration, we analyze the effect of rule of law on public support for democracy, finding that rule of law has a positive or negative effect on support for democracy depending on the choice of the indicator. Our work thus contributes to research on the sensitivity of descriptive and causal inference to the choice of cross-national indicators (Mudde and Schedler 2010).

Rule of law indicators

We examine four main sources of cross-national indicators of the rule of law used in the social sciences, selected based on their established reputation, availability of indicators of

the rule of law with interval-scaled measures, and broad country coverage dating back at least to 1990.[1] These criteria exclude a number of data sets, such as the ordinal measures from Freedom House and Polity IV and most of the other data sets with rule of law indicators discussed by Skaaning (2009) and Møller and Skaaning (2011). In this section, we describe the selected rule of law indicators, the operationalizations of the rule of law they rely on, the source data and aggregation strategies. Given that there is not a single accepted definition of the rule of law (cf. Skaaning 2009), we abstain from discussing the different theoretical approaches and their correspondence to the indicators, and focus on comparing the scope of the indicators themselves.

The Varieties of Democracy project's (VDem) *Rule of Law index* measures the extent to which 'laws [are] transparently, independently, predictably, impartially, and equally enforced, and to what extent do the actions of government officials comply with the law' and is composed of 15 indicators of compliance with the high court and judiciary, court independence, respect of the constitution by the executive, impartiality of public administration, law transparency, access to justice, judicial accountability, and judicial and public sector corruption (Coppedge et al. 2019a, 269). Each of the 15 indicators corresponds to questions answered on fully labelled ordinal scales by country experts, with the results aggregated by Bayesian factor analysis models (Coppedge et al. 2019a, 2019b).

The International Institute for Democracy and Electoral Assistance (IDEA) has recently created the Global State of Democracy data set that provides measures of *Impartial Administration* – one of five attributes of democracy, which 'concerns how fairly and predictably political decisions are implemented' (IDEA 2019, 248). It includes two subattributes: *Absence of Corruption* and *Predictable Enforcement*, which map onto the two main components of the rule of law from the VDem data. The *Impartial administration* measure consists of seven out of the VDem indicators included in the rule of law index described above, as well as two indicators from the International Country Risk Guide on corruption and bureaucratic quality (Tufis 2017). Thus, IDEA's rule of law index substantially overlaps with VDem's.

World Bank's WGI include a *Rule of Law* dimension (Kaufmann, Kraay, and Mastruzzi 2011) defined as capturing

> perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. (World Bank 2019)

It relies on around 200 indicators from about 30 sources, such as the Freedom House, the World Economic Forum and cross-national surveys. The data sources also include the Varieties of Democracy *Liberal component index*, which itself is a composite index including some of the elements of the VDem rule of law index described earlier, in addition to indicators of individual liberties and legislative constraints on the executive. The publicly available data provide raw values of separate variables from most of the data sources.

Democracy Barometer's (DB) *Rule of Law* index is a component of the *Freedom* dimension of the quality of democracy, and includes two sub-components: equality before the law and the quality of the legal system (Merkel et al. 2018a, 2018b). Equality before the law is measured with indicators of constitutional provisions for impartial courts, effective independence of the judiciary and effective impartiality of the legal system. Quality of the legal system consists of constitutional provisions for judicial

professionalism, as well as confidence in the justice system and in the police. For each of the six areas (three per sub-component) there are two indicators, which combine data from different sources. For example, confidence in the legal system, part of the 'confidence in the justice system' indicator, which is part of the 'quality of the legal system' sub-component of the rule of law index, combines estimates of average levels of confidence in the legal system from nine cross-national survey projects, including some of the ones used to construct the WGI index. Other data sources include, for example, the Global Competitiveness Report and the Bertelsmann Transformation Index, which are also used by WGI. Altogether, the documentation lists 11 data sources used to construct the rule of law index. The publicly available data only include values for the 12 indicators, not the most disaggregated variables from the various data sources.

Table 1 summarizes the main characteristics of the four rule of law indicators. A more detailed breakdown of the source indicators is available in the On-line Appendix. Based on this summary, the indicators from the VDem and IDEA project seem to be the most similar both in terms of definitions and measures, while the WGI indicators look the most distinct, due to its inclusion of property rights, crime and violence. Still, all indicators ostensibly measure the same concept, as also suggested by their names, so despite some differences in definitions and data sources, they can be reasonably expected to reflect the same true degree of respect for the rule of law in the given country and year.

The use of rule of law indicators in published research

To examine how the indicators are used in substantive social science research, we searched the archives of seven reputable political science journals from the last five years to identify papers focusing on the rule of law.[2] After screening the papers returned in the search, we identified 10 papers whose analysis (or one of the analyses) was quantitative and used one of the rule of law indicators from multiple countries. All papers use country-years as units of analysis and most employed regression models that accounted for the repeated measures within countries by standard error corrections or by including country fixed effects or multilevel modelling. Rule of law was a dependent variable in six

Table 1. Comparison of rule of law indicators.

Indicator	Dimensions	Data source
VDem	Transparent, independent, predictable, impartial and equal enforcement of laws	Own expert surveys
	Compliance of government officials with the law	
IDEA	Absence of public authority corruption	Expert surveys from VDem and ICRG
	Predictable enforcement of public authority	
WGI	Perceptions of the extent to which agents have confidence in and abide by the rules of society	Multiple existing data sources, including surveys of experts, firms and general populations, synthet ic indicators
	Quality of contract enforcement	
	Property rights	
	Quality of the courts and the judicial system	
	Quality of the police	
	Likelihood of crime and violence	
DB	Equality before the law	Multiple existing data sources, including surveys of experts, firms and general populations, synthetic indicators
	Quality of the legal system	

Note: VDem = Varieties of democracy, IDEA = International Institute for Democracy and Electoral Assistance, WGI = Worldwide Governance Indicators, DB = Democracy Barometer, ICRG = International Country Risk Guide.

papers and an independent variable in the other four. Summary information about the analyzed papers is provided in the On-line Appendix.

Among the 10 journal papers, 7 papers used the WGI indicator, 2 papers used the VDem rule of law indicator and 1 paper used the DB indicator. None of the papers we analyzed used the IDEA rule of law indicator, which may be due to the relatively recent publication of the IDEA dataset in 2019. While the WGI have been criticized for their questionable validity, our review shows that they remain an important data source for academic researchers (Thomas 2010; Langbein and Knack 2010; cf. Kaufmann, Kraay, and Mastruzzi 2010).

Only 1 of the 10 papers discussed different applicable rule of law indicators, chose 2 of them for the main analysis and included the results with the remaining 2 indicators in supplementary materials. Another paper used the WGI to measure the rule of law, and a different data source to measure an alternative concept, the quality of the criminal justice system. The remaining papers justified the choice of the chosen indicators primarily by referring to their wide use and broad coverage (cf. Møller and Skaaning 2011) or provided no justification at all. None of the papers we examined mentioned or took into account in models the uncertainty estimates provided in the VDem and WGI data.

In sum, most of the papers we analyzed provide no evidence that the choice of the rule of law indicator was informed by theoretical considerations or intended to maximize the match between the indicator's operationalization with the meaning and aspect of rule of law relevant for the given analysis, or that the authors were aware of the differences in definitions, source data and aggregation approaches used by different projects. We conclude that scholars tend to treat rule of law indicators from different projects as interchangeable and that the practice of examining the sensitivity of empirical results to the choice of the indicators is still rare (cf. Mudde and Schedler 2010). In other words, despite the broad offer, the competition between different rule of law indicators remains limited, and the potential of the 'marketplace of indicators' is not being exploited. In the next sections, we examine to what extent the assumption of interchangeability of the rule of law indicators is warranted in descriptive and causal analyses.

Consistency between indicators

The rule of law indicators we analyze provide data for 62 countries i over years t (the number of years per country varies between 14 and 21). x_{it} and y_{it} are values of two different rule of law indicators from country i and year t, and $\overline{x}_i$ and $\overline{y}_i$ are the average values of the two indicators, x and y, respectively, within country i across all years t. To check if there is cross-indicator agreement in the variation of rule of law, we first calculate correlations for pairs of indicators in the pooled data set, i.e. $cor(x_{it}, y_{it})$. We find uniformly high consistency with correlation coefficients in the range of around 0.8 and 0.9, as

Table 2. Correlations between pairs of rule of law indicators.

	Overall			Between countries			Average within counties		
	DB	WGI	IDEA	DB	WGI	IDEA	DB	WGI	IDEA
VDem	0.79	0.89	0.91	0.83	0.91	0.91	0.04	0.14	0.62
DB		0.92	0.86		0.95	0.89		0.19	0.11
WGI			0.93			0.95			0.04

demonstrated on the left-hand side of Table 2. Correlations between country means of the selected indicators, that is, correlations between averages within each country over all years for which data are available, $cor(\overline{x_i}, \overline{y_i})$, are higher yet, and in the analyzed example reach 0.95 (see the centre of Table 2). It is thus clear that high overall correlations are driven by between-country consistency in measurement. The right-hand side part of Table 2 presents averages of within-country correlations across all countries expressed as $\frac{1}{n}\sum_{c=1}^{n} cor(x_{i=c,\,t}\ y_{i=c,\,t})$. Average within-country correlations are much lower, and range from roughly 0 to 0.2, with the exception of the pair of indicators that are largely based on the same data, i.e. VDem and IDEA.

Figure 1 presents distributions of within-country correlations between all pairs of indicators. Consistent with the high average within-country correlation, VDem and IDEA indicators are positively correlated in most countries, but even there, for some countries the correlation is minimal or negative. For the remaining pairs of indicators, the correlations are negative in large proportions of countries.

Table 3 summarizes within-country correlations in a different way. It shows, for each pair of indicators, the proportions of countries, in which correlations are significantly negative (at the customary 5% level), not statistically significantly different from zero, and significantly positive. Consistent with Figure 1, rule of law indicators from the VDem and IDEA projects have the highest share of significant positive correlations, 74%. Correlations among the remaining pairs of indicators are significantly positive in about a fourth to a third of countries. For each indicator pair, there exist at least a couple of countries, in which the correlation is significant and negative. The remaining cases, which for most indicator pairs account for over half of the analyzed countries, are countries where the respective correlations are too low to reach statistical significance.

Negative correlations may result from situations where both indicators change only slightly, but in opposite directions. Such instances would still be problematic for models, but easier to understand. Figure 2 takes a closer look at rule of law trajectories in four countries: Austria, Greece, Poland and Serbia (plots for all countries are available in the Appendix). These countries were selected because of the overall low or negative correlations among pairs of rule of law indicators and with the goal of representing countries with low, middle and high rule of law scores. In Austria, rule of law since 2010 deteriorated
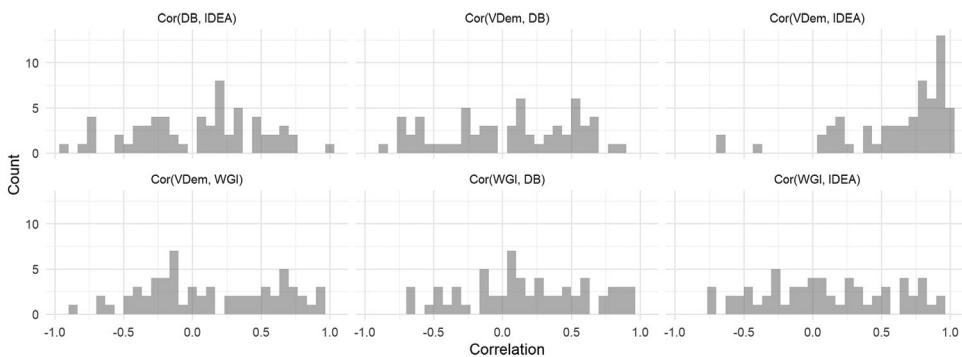

Figure 1. Distributions of within-country correlations between pairs of democracy indicators.

Table 3. Within-country correlations between pairs of rule of law indicators.

Pair		Negative, significant*	Not significantly different from 0	Positive, significant
DB, IDEA	N	9	39	14
	Proportion	0.15	0.63	0.23
VDem, DB	N	13	32	17
	Proportion	0.21	0.52	0.27
VDem, IDEA	N	2	14	46
	Proportion	0.03	0.23	0.74
VDem, WGI	N	6	36	20
	Proportion	0.1	0.58	0.32
WGI, DB	N	5	37	20
	Proportion	0.08	0.6	0.32
WGI, IDEA	N	10	35	17
	Proportion	0.16	0.56	0.27

* Significantly different from 0 at the 5% level.

according to DB but improved according to IDEA. Greece saw declines in the WGI and DB indicators since the late 2000s, but the other two indicators remained rather stable or even improved slightly. In Poland, rule of law was relatively stable according to VDem, declined up to 2005 and improved since according to WGI, and experienced ups and downs according to IDEA and DB. All indicators are in agreement, however, about the decline since 2015. In Serbia, WGI shows a long-term increasing trend in the rule of law, VDem and IDEA suggest stability, and DB dipped around 2006 as WGI noted a sizeable increase.

Cointegration tests are an alternative to correlations for establishing co-variation in time-series analyses. Most briefly, two time series are said to be cointegrated, if neither of the time series is stationary, but their linear combination is stationary (Granger 1981). As already mentioned, correlation picks up on the smallest changes, which in the case of time series may lead to counterintuitive results, as illustrated e.g. by Damghani et al. (2012). Cointegration tests, on the other hand, are more robust to small changes in the indicators that may not be meaningful.

We ran cointegration tests for all pairs of rule of law indicators separately for each of the 62 countries included in the analysis.[3] For nine countries, none of the tests turned out statistically significant at the 0.05 level, indicating no evidence that any of the six pairs of indicators can be considered cointegrated. Only in two countries cointegration tests returned a significant result for all pairs of indicators. In the remaining countries some pairs tested as cointegrated and some did not.
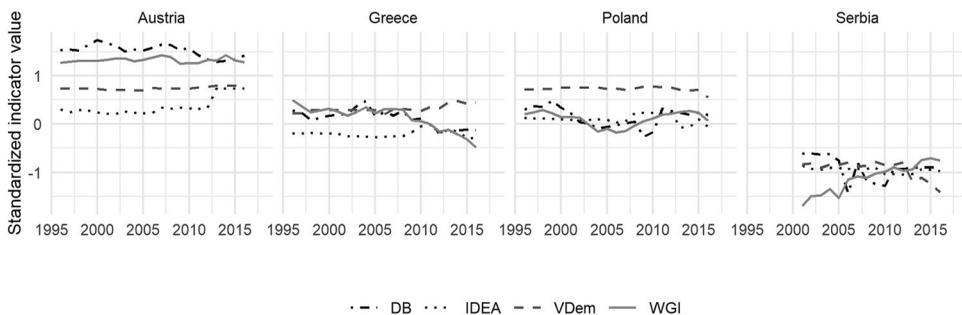
Figure 2. Rule of law trends according to different indicators for selected countries. Standardized values.

Rule of law and popular support for democracy

To illustrate the consequences of the inconsistencies in the measurement of rule of law, we examine the effect of the rule of law on popular support for democracy. There is a rich literature on the links between state performance including the quality of governance and societal political support, and discussing it is beyond the scope of this note. Most generally, one – and arguably dominant – line of reasoning, sometimes referred to as democratic learning theories, connects positive experiences with the state to both diffuse and specific political support (Bratton and Mattes 2001; Boräng, Nistotskaya, and Xezonakis 2017; Magalhães 2013; Park 2016). A contrasting approach argues for a thermostatic model of public opinion (Wlezien 1995), where an increase in the supply of democracy would lead to a decline in support for democracy, while deterioration of democracy would bring about increases in democratic support following the rule that 'one values what one does not have'. To date, few empirical studies test these competing claims with longitudinal data. The analysis by Claassen (2020) is a rare exception; it examines the effects of democracy on support for democracy relying on democracy indicators from the VDem project and measures of average levels of support for democracy based on an aggregation of data from multiple cross-national survey projects. The author finds support for the thermostatic hypothesis. There is little agreement in the literature regarding the aspects of governance, or 'system effectiveness' (Lipset 1959), to which citizens react most strongly, and the rule of law is considered one of the important factors. Thus, in our analysis, instead of broad indicators of democracy, we use indicators of the rule of law from the four sources discussed earlier.

Country-year levels of support for democracy, the dependent variable in our analysis, come from Claassen's (2020) paper mentioned above. The time series in democratic support were estimated with Bayesian item response theory models based on survey variables measuring democratic preferences from almost 1400 national surveys from 14 cross-national survey projects (Claassen 2019a, see 2019b for a detailed description of the method). Support for democracy is estimated using all available survey items on support for democracy from all available cross-national surveys, so there is no obvious alternative way of measuring it directly.

As independent variables, we use rule of law indicators from VDem, DB, IDEA and WGI datasets, standardized within the country-year sample. We control for GDP *per capita* to capture economic performance – the key competitor of governance in theoretical explanations of political support. We restrict the analysis to European countries, where all indicators have sufficient country coverage, and we use the set of countries and years available for all four rule of law indicators to improve the comparability of the results. We note however, that the common practice for published analyses is to use data for as many countries and years as available for a single chosen rule of law indicator (and other variables).

To model the effects of rule of law on support for democracy, we estimate two-way (country and year) fixed effects panel regression models. The model is specified as follows:

$$Y_{it} = \alpha_i + \gamma_t + \beta_1 X_{it} + \beta_2 Z_{it} + \epsilon_{it} \qquad (1)$$

where X_{it}, Y_{it} and Z_{it} represent rule of law, support for democracy, and GDP *per capita*, respectively, α_i are country fixed effects, γ_t are year fixed effects, and β_1 represents the

Table 4. Results of two-way fixed effects panel regression of rule of law on support for democracy.

Coef.	VDem	IDEA	WGI	DB
Rule of law	−.231 (.047)*	−.309 (.066)*	−.144 (.071)*	.245 (.047)*
GDP per capita	.065 (.097)	.120 (.101)	.113 (.119)	.018 (.095)
R^2	.947	.946	.945	.947
Num. obs.	629	629	629	629

Note: Standard errors are depicted in brackets. * = Coefficient is significantly different from zero at the 5% level. R^2 = Proportion of variance explained. Num. obs. = Number of observations.

effect of the rule of law on support for democracy, which is of main interest in this analysis.

The analysis was performed in R (R Core Team 2020) with the plm package (Millo 2017). We also used tidyverse (Wickham et al. 2019) for manipulating data, countrycode (Arel-Bundock, Enevoldsen, and Yetman 2018) for switching between country codes, knitr (Xie 2020) for creating reproducible documents, kableExtra (Zhu 2019) for creating complex tables, ggplot2 (Wickham 2016) for making plots and texreg (Leifeld 2013) for formatting model output.

As Table 4 shows, the results are mixed. Out of the four models, three suggest a negative effect of the rule of law on support for democracy. The fourth model, using the DB indicator, suggests a positive effect of similar magnitude to average of the negative coefficients of the other three models. Our earlier comparison of the definitions and source data used to construct the different indicators led us to conclude that the WGI indicator would be the most distinct from among the four, yet it is the DB indicator that suggests the opposite direction of the effect on support for democracy than the other three. The analysis of within-country correlations presented in Table 2, on the other hand, shows a similarly low average correlation between the VDem and DB indicators as between the WGI and IDEA indicators. Meanwhile, in the regression analysis WGI and IDEA point to effects in the same direction, and VDem and DB suggest effects in opposite directions. This is all to show that the effect of the choice of indicators on results of multi-variable analyses is difficult to predict based on a review of the documentation or exploration of the indicators.

The ambiguity regarding the association between rule of law and support for democracy becomes apparent when performing analyses using different rule of law indicators. Typically, each analysis uses only one indicator and draws conclusions based on that model's results. As our example shows, the risk of overconfidence in these results and conclusions is very high.

Conclusion

In this note, we analyzed the longitudinal correspondence between well-regarded indicators of the rule of law, an important aspect of the quality of governance, finding substantial differences in within-country trajectories across different indicators. In the most drastic but not so rare cases, within-country correlations are negative: when rule of law in a given country improves according to one indicator, it deteriorates according to another. Cointegration tests support the claim about substantial differences between country time series of different rule of law indicators. Of course, such differences could be expected to influence the results of statistical analyses that model changes in the rule of law. Indeed, our analysis of the effects of rule of law on support for democracy

yielded results leading to different conclusions depending on which indicator was used. The association would be interpreted as positive when using the DB data and negative when using data from the VDem, IDEA or WGI data sets.

The problem extends beyond this particular set of indices and at minimum applies to indicators of rule of law, democracy and corruption in the four data sources we described in this note. For any longitudinal analysis with one of these indicators, an analogous analysis with a different indicator will likely yield a different – sometimes opposite – result. This situation is especially problematic when models are used to adjudicate between competing theoretical explanations that predict opposite effects. The inconsistencies in governance indicators may also be consequential in studies that use changes in governance as a criterion for the selection of countries into the analysis, or analyses that discretize continuous measures to identify political transitions.

It seems that these over-time inconsistencies in governance indicators have so far gone underappreciated. Of the data sources described in this note, only WGI has come under scrutiny for its validity (Thomas 2010; Langbein and Knack 2010; cf. Kaufmann, Kraay, and Mastruzzi 2010), although some authors have also pointed to potential inaccuracies in VDem (Bakke and Sitter 2020, fn. 5). Lueders and Lust (2018) discuss disagreements in measures of regime change, and Cope, Crabtree, and Fariss (2019) point to differences in indicators of state repression. Documenting similar discrepancies in the present study suggests that the problem is widespread and affects much of comparative social science research, thus contributing to the fragmentation of research and inhibiting its cumulative development.

Researchers using country-level measures should be aware of the available indicators, and – especially when conducting longitudinal analyses – carefully consider their choice taking into account the theoretical framework and conceptualizations guiding their study, as well as the definitions, source data and aggregation techniques used to construct the indicators. If more than one indicator seems appropriate, the most suitable one could be identified by examining country trends as part of 'case oriented validation' (Adcock et al. 2001), by asking whether changes in the indicators are consistent with country expertise. A close examination would necessitate access to all the raw source data, including individual expert ratings and components of the indexes used by DB, IDEA and WGI.[4] The framework for assessing the quality of measures of democracy by Pickel, Stark, and Breustedt (2015) or the strategy of validation proposed by McMann et al. (2016) could help in this process. Alternatively, researchers may proceed with analyses and test the robustness of their claims against different indicators. In some cases, the discrepancies in rule of law and other indicators will likely translate into conflicting results. Returning to the conceptualization or validation stage could resolve some of the conflicts. Even if not, from the scientific point of view, studies that point to disagreement in the results of analyses with different indicators would be preferable to the current practice of publications relying on a single indicator without sufficient justification.

As a rule, applied researchers should also consider the uncertainty that results from the estimation of country-year levels of VDem's, WGI's, and IDEA's indicators and incorporate it into analyses rather than using only point estimates.[5] Many measures of social phenomena are products of complex procedures including defining the construct, operationalization and quantification, each requiring multiple decisions. The memory of these steps and decisions is lost in the moment when indicators are entered into models as if they were

observed values measured without error. Incorporating uncertainty into analyses would potentially reduce the number of very small but statistically significant (and hence commonly interpreted) effects in the empirical literature on the subject. It could also inspire a discussion about the desirable level of precision of governance indicators, as well as motivate research on the propagation of uncertainty in synthetic indicators that result from multiple stages of aggregation.

Notes

1. Our findings apply equally to other aspects of governance, including corruption and broad measures of democracy, for which also multiple country-year indicators are available (cf. On-line Appendix).
2. The search query in the Web of Science Core Collection database was as follows: ((SO=(AMERICAN POLITICAL SCIENCE REVIEW) OR SO=(BRITISH JOURNAL OF POLITICAL SCIENCE) OR SO=(DEMOCRATIZATION) OR SO=(EUROPEAN JOURNAL OF POLITICAL RESEARCH) OR SO=(EUROPEAN POLITICAL SCIENCE REVIEW) OR SO=(INTERNATIONAL POLITICAL SCIENCE REVIEW) OR SO=(JOURNAL OF EUROPEAN PUBLIC POLICY))) AND PY=(2016 OR 2017 OR 2018 OR 2019 OR 2020 OR 2021) AND AB=("RULE OF LAW").
3. We used the aTSA package (Qiu 2015) in R (R Core Team 2020) to check the order of integration of the time series and to conduct the cointegration tests.
4. A possible conclusion of such validation could be that the source data used to construct rule of law and other governance indicators are too noisy to accurately gauge year-on-year changes, as these changes are in most cases very small compared to cross-country differences, for evaluating which the indicators where originally intended. An analysis of separate components of the four rule of law indicators (available in the On-line Appendix) shows that, when pooling data across countries and years, correlations among them – even between components of different rule of law indicators – are typically very high. Average within-country correlations are much lower, and sometimes very low, even among components of the same indicator. A separate issue are potential limitations to longitudinal comparability even within the same dataset, e.g. due to changes in the availability of the source data used to construct the WGI (Kaufmann, Kraay, and Mastruzzi 2011).
5. It is worth noting that uncertainty estimates provided in the WGI and IDEA data capture only part of the overall uncertainty involved in the construction of these indicators, namely the part related to the aggregation of the source data, while ignoring, for example, the uncertainty stemming from the aggregation of data from general population surveys used in WGI, for the uncertainty provided for the VDem indicators, which both WGI and IDEA use as source data. Attempts to re-estimate overall uncertainty of rule of law (and other governance indicators), taking into account all its sources, would require to all source data at the most disaggregated level, which are currently not publicly available. Further analyses could examine whether uncertainty intervals for rule of law indicators are higher in countries with higher disagreements in the rule of law ratings. The data that is publicly available at present does not enable such analysis.

Acknowledgement

Data availability statement

The on-line appendix, data and replication materials are available at https://osf.io/wsy9v/.

Disclosure statement

Funding

ORCID

Marta Kołczyńska http://orcid.org/0000-0003-4981-0437
Paul-Christian Bürkner http://orcid.org/0000-0001-5765-8995

References

Adcock, R., and D. Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–546. doi:10.1017/S0003055401003100.

Arel-Bundock, V., N. Enevoldsen, and C. J. Yetman. 2018. "Countrycode: An R Package to Convert Country Names and Country Codes." *Journal of Open Source Software* 3 (28): 848. doi:10.21105/joss.00848.

Bakke, E., and N. Sitter. 2020. "The EU's Enfants Terribles : Democratic Backsliding in Central Europe Since 2010." *Perspectives on Politics* 1–16. doi:10.1017/S1537592720001292.

Boese, V. A. 2019. "How (not) to Measure Democracy." *International Area Studies Review* 22 (2): 95–127. doi:10.1177/2233865918815571.

Bollen, K. A. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 45 (3): 370. doi:10.2307/2095172.

Boräng, F., M. Nistotskaya, and G. Xezonakis. 2017. "The Quality of Government Determinants of Support for Democracy." *Journal of Public Affairs* 17 (1–2): e1643. doi:10.1002/pa.1643.

Bratton, M., and R. Mattes. 2001. "Support for Democracy in Africa: Intrinsic or Instrumental?" *British Journal of Political Science* 31 (03): 447–474. doi:10.1017/S0007123401000175.

Claassen, C. 2019a. "Replication Data for: In the Mood for Democracy? Democratic Support as Thermostatic Opinion." *Harvard Dataverse*, doi:10.7910/DVN/FECIO3.

Claassen, C. 2019b. "Estimating Smooth Country–Year Panels of Public Opinion." *Political Analysis* 27 (1): 1–20. doi:10.1017/pan.2018.32.

Claassen, C. 2020. "In the Mood for Democracy? Democratic Support as Thermostatic Opinion." *American Political Science Review* 114 (1): 36–53. doi:10.1017/S0003055419000558.

Cope, K. L., C. Crabtree, and C. J. Fariss. 2019. "Patterns of Disagreement in Indicators of State Repression." *Political Science Research and Methods* 8 (1): 178–187. doi:10.1017/psrm.2018.62.

Coppedge, M., J. Gerring, C. H. Knutsen, S. I. Lindberg, J. Teorell, D. Altman, M. Bernhard, et al. 2019a. "V-Dem Codebook v9." April. Varieties of Democracy (V-Dem) Project. https://www.v-dem.net/media/filer_public/e6/d2/e6d27595-9d69-4312-b09f-63d2a0a65df2/v-dem_codebook_v9.pdf.

Coppedge, M., J. Gerring, C. H. Knutsen, S. I. Lindberg, J. Teorell, D. Altman, M. Bernhard, et al. 2019b. "V-Dem Dataset v9." Varieties of Democracy (V-Dem) Project. https://doi.org/10.23696/vdemcy19.

Damghani, B. M., D. Welch, C. O'Malley, and S. Knights. 2012. "The Misleading Value of Measured Correlation." *Wilmott* 2012 (62): 64–73. doi:10.1002/wilm.10167.

Granger, C. W. J. 1981. "Some Properties of Time Series Data and Their Use in Econometric Model Specification." *Journal of Econometrics* 16 (1): 121–130. doi:10.1016/0304-4076(81)90079-8.

Högström, J. 2012. "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. doi:10.1017/gov.2012.10.

IDEA. 2019. *The Global State of Democracy 2019: Addressing the Ills, Reviving the Promise*. Stockholm: International Institute for Democracy; Electoral Assistance. doi:10.31752/idea.2019.31.

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2010. "Response to 'What Do the Worldwide Governance Indicators Measure?'" *European Journal of Development Research* 22 (1): 55–58. doi:10.1057/ejdr.2009.49.

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2011. "The Worldwide Governance Indicators: Methodology and Analytical Issues." *Hague Journal on the Rule of Law* 3: 220–246.

Langbein, L., and S. Knack. 2010. "The Worldwide Governance Indicators: Six, One, or None?" *Journal of Development Studies* 46 (2): 350–370. doi:10.1080/00220380902952399.

Leifeld, P. 2013. "texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables." *Journal of Statistical Software* 55 (8): 1–24. doi:10.18637/jss.v055.i08.

Lipset, S. M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53 (1): 69–105. doi:10.2307/1951731.

Lueders, H., and E. Lust. 2018. "Multiple Measurements, Elusive Agreement, and Unstable Outcomes in the Study of Regime Change." *The Journal of Politics* 80 (2): 736–741. doi:10.1086/696864.

Lührmann, A., K. L. Marquardt, and V. Mechkova. 2020. "Constraining Governments: New Indices of Vertical, Horizontal, and Diagonal Accountability." *American Political Science Review*, 1–10. doi:10.1017/S0003055420000222.

Magalhães, P. C. 2013. "Government Effectiveness and Support for Democracy." *European Journal of Political Research* 53 (1): 77–97. doi:10.1111/1475-6765.12024.

Malito, D. V. 2014. "Measuring Corruption Indicators and Indices." *SSRN Electronic Journal*, doi:10.2139/ssrn.2393335.

McMann, K. M., D. Pemstein, B. Seim, J. Teorell, and S. I. Lindberg. 2016. "Strategies of Validation: Assessing the Varieties of Democracy Corruption Data." *SSRN Electronic Journal.*, doi:10.2139/ssrn.2727595.

Merkel, W., D. Bochsler, K. Bousbah, M. Bühlmann, H. Giebler, M. Hänni, L. Heyne, et al. 2018a. *Democracy Barometer. Codebook. Version 6*. Aarau: Zentrum für Demokratie.

Merkel, W., D. Bochsler, K. Bousbah, M. Bühlmann, H. Giebler, M. Hänni, L. Heyne, et al. 2018b. *Democracy Barometer. Methodology. Version 6*. Aarau: Zentrum für Demokratie.

Millo, G. 2017. "Robust Standard Error Estimators for Panel Models: A Unifying Approach." *Journal of Statistical Software* 82 (3): 1–27. doi:10.18637/jss.v082.i03.

Møller, J., and S.-E. Skaaning. 2011. "On the Limited Interchangeability of Rule of Law Measures." *European Political Science Review* 3 (3): 371–394. doi:10.1017/s1755773910000421.

Mudde, C., and A. Schedler. 2010. "Introduction: Rational Data Choice." *Political Research Quarterly* 63 (2): 410–416. doi:10.1177/1065912910371286.

Park, C.-M. 2016. "Quality of Governance and Regime Support." *Asian Journal of Comparative Politics* 2 (2): 154–175. doi:10.1177/2057891116675769.

Pickel, S., T. Stark, and W. Breustedt. 2015. "Assessing the Quality of Quality Measures of Democracy: A Theoretical Framework and its Empirical Application." *European Political Science* 14 (4): 496–520. doi:10.1057/eps.2015.61.

Qiu, D. 2015. *ATSA: Alternative Time Series Analysis*. https://CRAN.R-project.org/package=aTSA.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/.

Skaaning, S.-E. 2009. "Measuring the Rule of Law." *Political Research Quarterly* 63 (2): 449–460. doi:10.1177/1065912909346745.

Standaert, S. 2015. "Divining the Level of Corruption: A Bayesian State-Space Approach." *Journal of Comparative Economics* 43 (3): 782–803. doi:10.1016/j.jce.2014.05.007.

Thomas, M. A. 2010. "What Do the Worldwide Governance Indicators Measure?" *The European Journal of Development Research* 22 (1): 31–54. doi:10.1057/ejdr.2009.32.

Tufis, C. D. 2017. *The Global State of Democracy Indices: Technical Procedures Guide*. Stockholm: International Institute for Democracy; Electoral Assistance. doi:10.31752/idea.2018.74.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. http://ggplot2.org.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4: 1686. doi:10.21105/joss.01686.

Wlezien, C. 1995. "The Public as Thermostat: Dynamics of Preferences for Spending." *American Journal of Political Science* 39 (4): 981–1000.

World Bank. 2019. "Rule of Law: Number of Sources." https://databank.worldbank.org/metadataglossary/worldwide-governance-indicators/series/RL.EST.

Xie, Y. 2020. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://yihui.org/knitr/.

Zhu, H. 2019. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.Rproject.org/package=kableExtra.

Appendix

Rule of law: all country trends
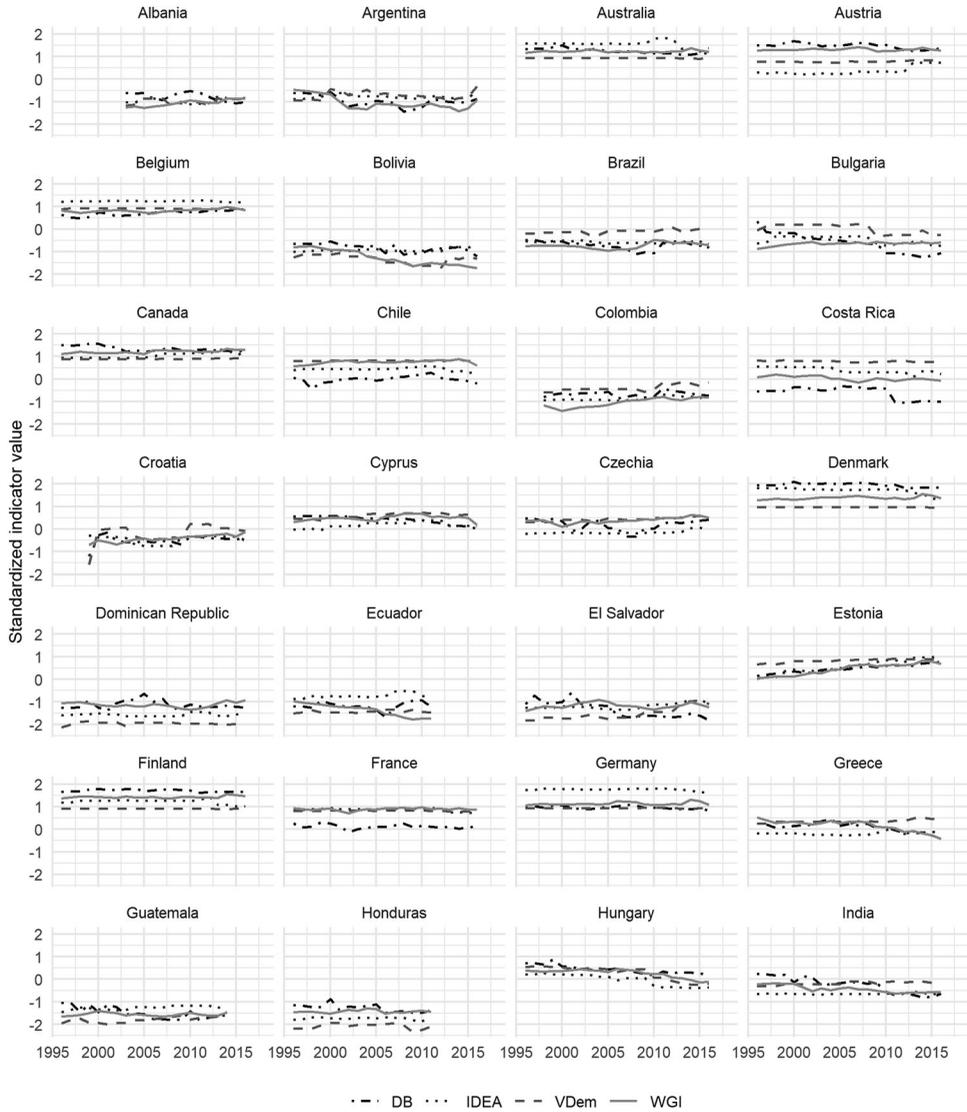
Figure A1. Rule of law trends according to different indicators for selected countries (Albania – India).

Figure A2. Rule of law trends according to different indicators for selected countries (Ireland – Venezuela).