On the limited interchangeability of rule of law measures

J Ø R G E N   M Ø L L E R * A N D   S V E N D - E R I K   S K A A N I N G

Associate Professor, PhD, Department of Political Science, Aarhus University, Bartholins Allé, Århus C, Denmark

During the latest decade, empirical research on the causes and consequences of the rule of law has expanded and, in the process, become extremely influential. However, we show that a number of widely used indices of the rule of law are not interchangeable. This lack of interchangeability is reflected in the fact that they are based on different defining attributes, to some extent cover distinct empirical scopes, do not correlate highly with each other, and support different explanatory factors. Until a consensus has been established with respect to the conceptualization of the rule of law, scholars are thus not free to opt for the measure that fits their data requirements best regarding spatial and/or temporal scope. Instead, they must carefully assess the content validity vis-à-vis their stipulated definition of the rule of law. Given the amount of money and time poured into the rule of law agenda, the problems identified reflect the lack of maturity of 'good governance' research.

Keywords: rule of law; indices; interchangeability; missingness; content validity

Introduction

Introducing a recent symposium on data use in comparative politics, Mudde and Schedler (2010: 411) dishearteningly note that though we have recently witnessed

> vigorous and sophisticated debates on the quality of cross-national data … we still do not see evaluations of data quality appear in academic journals on a regular basis, and uncertainty over the quality of the cross-national data we use continues to be pervasive in many areas of research.

A number of influential articles and books have indeed targeted the quality of cross-national data sets (e.g. Adcock and Collier, 2001; Munck and Verkuilen, 2002; Goertz, 2005). Systematic templates for assessing and guiding both the data generating process and the content validity of the data do exist. Arguably, however, the proposed guidelines have been disseminated to the actual practices of those analyzing comparative politics only to a limited extent (Munck, 2009, 2010). One of the most striking examples is to be found in the rule of law research agenda.

* E-mails: jm@ps.au.dk, skaaning@ps.au.dk

371

Over the latest decades, the rule of law has become one of the most celebrated concepts within politics and academia alike. At times, the underlying phenomenon is construed as something close to a panacea for developing countries – politically as well as economically (Tamanaha, 2004: 1–6; The Economist, 2008). Yet, paradoxically, the interest in the rule of law has in no way been accompanied by a consensus as to how the concept should be defined (Tamanaha, 2004; Belton, 2005; Møller and Skaaning, 2010; Skaaning, 2010).

Consequently, scholars analyze very different things while maintaining that they deal with the rule of law. This is probably one reason why the actual empirical analyses of the causes and consequences of the rule of law have produced strikingly dissimilar results, depending on which rule of law measure has been used (see, e.g. Knack, 1996; Norton, 1998; Barro, 2000; Ali, 2003; Andrews and Montinola, 2004; Joireman, 2004; Butkiewicz and Yanikkaya, 2006). In this paper, we pursue this observation systematically and demonstrate that the present lack of equivalence in cross-national studies on the rule of law mirrors the fact that the most influential rule of law measures are not interchangeable.

This is probably the right place to note that we refrain from assessing the measurement validity (Adcock and Collier, 2001) and 'concept-measure consistency' (Goertz, 2005) of the individual indices, as such exercises basically presuppose what is missing, namely a widespread consensus about the definition of the rule of law. Instead, we heed Mudde and Schedler's (2010: 412–413) call for carrying out two specific sets of assessments of the 'substantive implications' of using different data sets. First, we assess the sensitivity of descriptive inferences to data selection. Second, we assess the sensitivity of causal inferences to data selection. On this basis, we develop some more general analytical points about the 'good governance' research field and the problem of cumulativity in the Conclusions.

Posing the questions

More precisely, we pose four questions. First, do the dominant measures of the rule of law differ with respect to the conceptualization, that is, as regards the defining attributes? Second, do the measures correlate with each other? Third, do the measures lend support to the same explanatory factors derived from the literature on the rule of law? Finally, do non-random patterns of missing data introduce bias with respect to the investigated explanatory factors in one or more of the indices?

Before attempting to answer these questions, however, the present lack of consensus needs to be addressed. This is definitely not something unique to the rule of law research agenda. Most sub-fields in the social sciences are characterized by similar definitional controversies. To some extent, this seems to indicate that a definitional consensus is not to be expected. What is more, it is not a problem insofar as the scholarly community is conscious about the lack of interchangeability. If scholars carefully opt for the rule of law index which best

accords with their stipulated definition of the rule of law *and* if their conclusions clearly state what has in fact been explained or used as an explanatory variable, the problem of interchangeability disappears. Indeed, if analysts conscientiously choose their measures with reference to what Adcock and Collier (2001) term content validity, the present lack of consensus may be beneficial for research as more dimensions of the rule of law are scrutinized.

However, such conscientious treatment of indices is not the norm at present. Most researchers simply opt for the index that fits their data requirements best without discussing the content validity of these measures. Bolaky and Freund (2006: 23, fn. 16) thus choose the Law and Order Index (Political Risk Services (PRS) below) rather than that of the World Bank (World Governance Indicators (WGI) below) for the sole reason that the latter does not contain year-by-year data for the early 1990s and onwards. Burnside and Dollar (2004: 9) also select the Law and Order Index – as a proxy for institutional quality – because it is 'available for many developing countries going back to the 1980s'. Even more tellingly, when describing their preferred measure of the dependent variable (i.e. rule of law), Andrews and Montinola (2004: 72) merely note that 'PRS provides the most complete data for the period of interest'. Knack (2002: 12) captures the standard practice well in observing about the Law and Order Index that:

> Because of its much better cross-country coverage … the ICRG [PRS] indicators have been the most widely used governance indicators in the cross-country empirical literature on economic performance.

The list could be expanded by studies, which do not explicitly argue that coverage was the criterion for the choice of data set – but, rather, remain silent (and maybe agnostic) on the question of the criteria for selecting a particular measure, or bluntly refer to the fact that the measure employed has been used in previous studies.[1]

When data convenience rather than content validity becomes the norm, non-interchangeability is indeed problematic. This is further accentuated by the fact that most analyses refrain from using several indices to test the robustness of their findings, which means that the results hinge completely on the chosen index.[2] In addition, the problem is enhanced because even many of the indices scrutinized in this article – which are selected, among other criteria, due to their

---

[1] For instance, in a recent paper, Bleaney and Dimico (2008: 6) provide a very parsimonious argumentation for their selection of indicators to measure their dependent variables: 'To proxy institutions we use two alternative indices. The first one is an index of property rights protection … published by the Fraser Institute. The second measure, which is available for considerably more countries, is an index of the rule of law supplied by Kaufmann *et al.*'

[2] Exceptions do exist (e.g. Norton, 1998; Sunde *et al.*, 2008). Joireman (2001) uses the civil liberty scores provided by FH as an alternative measure of rule of law but neither FH itself nor the great majority of scholars consider this measure to solely capture the rule of law. Tellingly, Andrews and Montinola (2004) use it as an independent variable in their model trying to explain the strength of rule of law.

Table 1. Originator, name, and scope of rule of law measures, 2005

Originator	Data set	Measure	Countries
BTI	Bertelsmann Transformation Index	Rule of law	119
FH	Freedom of the World	Rule of law	192
PRS	International Country Risk Guide	Law and order	139
Kaufman *et al.* (WGI)	Worldwide Governance Indicators	Rule of law	210
FI	Economic Freedom of the World Data	Legal structure and security of property rights	141
HF	Index of Economic Freedom	Property rights	155
IF	Ibrahim Index	Safety and rule of law	53

BTI = Bertelsmann Transformation Index; FH = Freedom House; PRS = Political Risk Services; WGI = World Governance Indicators; FI = Fraser Institute; HF = Heritage Foundation; IF = Mo Ibrahim Foundation.

wide coverage – exclude different parts of the world, and tend to do so in non-random ways. Bearing this in mind, we set out to assess whether the dominant measures are interchangeable.

Selecting rule of law indices

Among readily available data sets, we have identified seven rule of law measures, based on three criteria. First, they target the actual (*de facto*) level of the rule of law, rather than formal (*de jure*) commitments. Second, they primarily build on experts' assessments in the form of standards-based data.[3] Third, with the exception of the Ibrahim Index (see below), the scope of the data covers the contemporary degree of rule of law in at least the majority of the countries of the world.

The seven measures are presented in Table 1. The first is the rule of law sub-index from the *Bertelsmann Transformation Index* (BTI; 2006), which covers almost all non-OECD countries with more than two million citizens. The second is Freedom House's (FH; 2006) sub-category scores for the rule of law found in the *Freedom in the World Survey*. The third is the law and order index provided by Political Risk Services in its *International Country Risk Guide* (2007). The fourth is the aggregate rule of law indicator constructed by Kaufman *et al.* (2007) in the context of the WGI, which is based on many different data sources, including most of the other measures mentioned. The fifth is the sub-category scores for legal structure and security of property rights included in Fraser

---

[3] As Hawken and Munck (2009) have demonstrated in an appraisal of indices on corruption, the class of the sources may have strong ramifications for the data.

Institute's (FI; 2009) *Economic Freedom of the World Data*. The sixth is the measure of property rights, which is part of the *Index of Economic Freedom* provided by the Heritage Foundation (HF; 2009) and the Wall Street Journal. The seventh – and final – measure is *the Ibrahim Index of African Governance* provided by the Mo Ibrahim Foundation (IF; 2010).

Generally, these indices cover different years and differ significantly with regard to the number of countries included. The information presented in the table applies to 2005, which is employed in the subsequent analyses because it is one of the very few years that all seven measures cover. Henceforth, we use the abbreviations of Table 1 to refer to the respective measures.

With the partial[4] exception of the Ibrahim index, the seven measures arguably suffer from what may be termed an 'OECD bias' as they measure the world from the point of view of the well-functioning and affluent OECD countries.[5] Critics have pointed to this on a number of occasions. The FH has thus been accused of having a right-wing bias and of overstating the level of freedom in 'US-friendly' countries (Chomsky and Herman, 1988; Bollen, 1992: 205; Barahona, 2007; Giannone, 2010). Likewise, the BTI has been blamed of being ahistorical and Eurocentric, which implies that post-colonial countries are not assessed properly (Koelble and Lipuma, 2008). A similar charge has been leveled against the World Bank, which has been accused of supporting Western interests in general, and US business interests specifically, as it has pushed for a neo-liberal agenda (Uvin, 2002; Moore, 2007). Finally, the measures of the HF and the FI have been said to have a libertarian bent (Card and Freeman, 2002: 3; cf. Ashby and Sobel, 2008: 332).

At the same time, and as demonstrated in the conceptual appraisal below, the indices measure starkly different things. One important insight of this paper is exactly that two more particular clusters of indices can be identified. Some indices (FH and BTI) seem to target what is best termed 'political constitutionalism', whereas others (PRS, HF, and FI) seem to measure 'public order'. Most strikingly, a conceptual slide occurs when the degree of 'property rights' is understood as a measure of the rule of law. At most, such rights constitute one sub-component of the more general concept. This also comes out in Table 2, which illustrates the differences in defining attributes among the seven measures. One might thus – somewhat paradoxically – simultaneously question whether the indices are too uniformly Eurocentric *and* whether they simply measure different things, such as distinct aspects of the overarching concept of good governance rather than the rule of law.

Neither of these two problems, however, affects the objective of this paper. As mentioned, our aim is not to take stock of the validity of the rule of law

---

[4] 'Partial' because the Ibrahim Index – though created as a counterweight to the extant 'Western' measures – is basically developed by Harvard scholars, is not based on an alternative 'non-Western' conception of the rule of law, and actually combines a number of existing data generated by, for example, the Economist Intelligence Unit and the World Bank. As such, the index bears comparison with the WGI.

[5] We thank an anonymous reviewer for calling attention to this point of contention.

Table 2. Conceptualization of rule of law measures

Measure	Attributes	
Rule of law (BTI)	Checks and balances	Working separation of powers Independent judiciary Legal or political penalties for power abusing officials
	Civil rights	Protection of civil liberties and possibility of seeking redress for violations
Rule of law (FH)	Independent judiciary	Judiciary not subject to interference Judges appointed and dismissed in a fair and unbiased manner Judges rule fairly and impartially Governmental authorities comply with judicial decisions which are enforced Private concerns comply with judicial decisions which are enforced
	Rule of law in civil and criminal matters	Protection of defendants' rights, including presumption of innocence Access to independent, competent legal counsel Fair, public, and timely hearing by competent, independent, and impartial tribunal Prosecutors independent of political control and influence Prosecutors independent of powerful private legal/illegal interests Effective and democratic control of law enforcement officials Law enforcement officials free from the influence of non-state actors
	Protection from political terror, unjustified imprisonment, exile, torture, and freedom from war and insurgencies	No arbitrary arrests and detentions and no fabrication or planting of evidence No beating of detainees or use of excessive force or torture Conditions in pretrial facilities and prisons humane Citizens have means of effective petition and redress Population not subjected to physical harm, forced removal, etc. due to conflict/war

Table 2. *(Continued)*

	Laws, policies, and practices guarantee equal treatment	Various distinct groups have equality before the law
		Violence against such groups not widespread and perpetrators brought to justice
		No legal/*de facto* discrimination in employment, housing, etc. against such groups
		Women enjoy full equality in law and practice
		Non-citizens enjoy basic human rights
Law and order (PRS)	Strength and impartiality of the legal system	
	Popular observance of the law	
Rule of law (WGI)	Agents have confidence in the rules of society	
	Agents abide by the rules of society	
	High quality contract enforcement	
	High quality police	
	High quality courts	
	Low likelihood of crime	
	Low likelihood of violence	
Legal structure and security of property rights (Fraser)	Judicial independence	
	Impartial courts	
	Protection of property rights	
	Military interference in rule of law and the political process	
	Integrity of the legal system	
	Legal enforcement of contracts	
	Regulatory restrictions on the sale of real property	
Property rights (Heritage)	Law protects private property rights	
	Government enforces property laws	
	Risk of expropriation	
	Independence of the judiciary	
	Corruption within the judiciary	
	Ability and businesses to enforce contracts	

Table 2. *(Continued)*

Measure	Attributes	
Safety and rule of law (IF)	Personal safety	Safety of the person
		Violent crime
		Social unrest
		Human trafficking
		Domestic political persecution
	Rule of law	Strength of judicial process
		Judicial independence
		Property rights
		Time taken to settle a contract dispute
		Orderly transfers of power
		UN sanctions
	Accountability and corruption	Transparency and corruption
		Accountability of public officials
		Corruption in government and public officials
		Prosecution of abuse of office
		Corruption
		Accountability, transparency, and corruption in rural areas
	National security	Domestic armed conflict
		Government involvement in armed conflict
		Battle deaths
		Civilian deaths from civilian-targeted violence
		Refugees from the country
		Internally displaced persons
		International tensions

BTI = Bertelsmann Transformation Index; FH = Freedom House; PRS = Political Risk Services; WGI = World Governance Indicators; IF = Mo Ibrahim Foundation.
Note: Where the data providers have not provided an explicit definition, we present the indicators.

measures – or governance measures in general for that matter (see, e.g. Arndt and Oman, 2006; Thomas, 2007; Williams and Siddique, 2008) – but to scrutinize their interchangeability. Fraser's and Heritage's property rights measures are, for instance, time and again used as indicators of the rule of law in the literature (cf. Munck, 2003; Ríos-Figueroa and Staton, 2008), which is why we include them in this appraisal. More generally, we have attempted to include the most dominant indices which are actually employed as proxies for the rule of law in contemporary analyses. Using Adcock and Collier's (2001) valuable distinction, we compare indices which are based on different 'systematized concepts' while agreeing on the 'background concept' (the rule of law). That we are not really comparing like with like is thus part of the problem which calls for assessment in the first place.

A preliminary conceptual appraisal

The literature to a large extent uses these indices as if they were interchangeable. The first question is if that is warranted conceptually? As Table 1 makes clear, only four of the indices (BTI, FH, IF, and WGI) actually employ the wording 'rule of law', the fifth (PRS) opting instead for the more ambiguous 'law and order', whereas the sixth (FI) and seventh (HF) focus on the legal structure and/or property rights. These differences in nomenclature indicate that the measures are not interchangeable even on the conceptual level. This tentative conclusion is borne out as soon as we descend to the level of the defining attributes, illustrated in Table 2.

As should be clear from this overview, little or no consensus concerning the conceptualization of the rule of law exists when we focus on the actual defining attributes underpinning the indices. The rule of law research agenda thus differs significantly from the related democratization research agenda. To illustrate this, it is pertinent to touch upon Casper and Tufis (2003), who demonstrate that three prominent measures of democracy are not genuinely interchangeable. However, Casper and Tufis (2003: 197) also show that all of these indices commence from Dahl's (1971) definition of polyarchy and, consequently, that they are highly correlated (correlation coefficients between 0.85 and 0.92). Not even the first of these two premises are fulfilled in our case. In brief, the rule of law research agenda has not settled for a particular systematized concept.

Correlations between the indices

What about the second premise of Casper and Tufis' (2003) analysis: that the indices are highly correlated?[6] Considering the lack of consensus on the systematized

---

[6] Claims for the validity of measures are commonly based on their high statistical correlations with other measures of the same phenomenon. However, strongly correlated measures may all share biases and errors (Forewaker and Krznaric, 2001: 5; Munck and Verkuilen, 2002: 29) so that high correlations do not signify a high level of validity. It is, nevertheless, important for the question of interchangeability to

concept, one would expect a much lower consistency of correlations in the case of rule of law measures. Such is indeed the case, as illustrated in Table 3.

The seven rule of law measures correlate in the range of 0.21 to 0.80 (Kendall's tau-b). More particularly, two observations can be made. First, the correlations between BTI and FH and among WGI, FI, and HF are very high. Second, FH, BTI, and PRS show a relatively low degree of co-variation with a number of the other measures.

These patterns also emerge in a factor analysis (principal component) that extracts two components (with Eigenvalues above 1, accounting for 65 and 20%, respectively, of the variation).[7] BTI, FH, and PRS obtain the lowest loadings on the first component. Moreover, PRS exhibits a higher positive loading on the second component, whereas the loadings on this component are negative, and moderately high, in the cases of BTI and FH.

The consistently high correlations between the WGI measure and all the other indices are hardly surprising, considering the fact that the latter make up important parts of the former (or vice versa, as it is to some extent the case with respect to FI). Yet, other than that, only the correlations between BTI and FH operate on the same level. Most striking are the low correlations between the PRS and FH, BTI (0.11 and 0.36, respectively) and between the BTI, FI, and HF (0.40 and 0.51, respectively).

This indicates that some of the indices do not measure the same empirical phenomenon. To be more exact, two or even three clusters seem to exist. First, the FH and BTI indices correlate strongly. Second, the same is the case for FI and HF. Finally, WGI and IF can be linked with both clusters for the simple reason that they subsume the other measures to a large extent.

To further probe these patterns, we follow Adcock and Collier's (2001: 540–541) recommendation to assess the correlation between the measures and measures of neighboring, yet distinct, concepts. The most obvious choice is a measure capturing a purely Schumpeterian definition of democracy, that is, one which only includes barebones electoral aspects.[8] We use the sub-component 'electoral self-determination' from the CIRI Human Rights Data set[9] to capture such a minimalist conception of democracy. The results are reported in the first row of Table 4.

Two things are of interest here. First, PRS is once again the odd one out as the correlation with electoral process is a mere 0.08. Second, as regards FH and BTI, the correlations are on a par with the equivalent correlations among the rule of law indices (except that between FH and BTI). This goes to show that the

scrutinize whether the very high correlation between different measures of democracy (cf. Coppedge and Reinicke, 1990: 61; Alvarez *et al.*, 1996: 18–21) is repeated with the rule of law measures.

[7] The IF is excluded from this analysis because of its low coverage.

[8] With a 'thicker' definition – for example, of liberal democracy – the rule of law would be included in the intention of the concept, meaning that the measure would, if valid, automatically correlate highly with rule of law measures.

[9] http://ciri.binghamton.edu/.

Table 3. Bivariate correlations between rule of law indices (2005)

	Rule of law (BTI)	Rule of law (FH)	Law and order (PRS)	Rule of law (WGI)	Legal structure and property rights (FI)	Property rights (HF)	Safety and rule of law (IF)
Rule of law (BTI)	1.00 (119)	0.80 (119)	0.11 (100)	0.55 (119)	0.40 (100)	0.51 (108)	0.68 (40)
Rule of law (FH)		1.00 (192)	0.36 (138)	0.64 (192)	0.55 (139)	0.63 (154)	0.59 (53)
Law and order (PRS)			1.00 (139)	0.61 (139)	0.66 (123)	0.47 (130)	0.35 (37)
Rule of law (WGI)				1.00 (210)	0.77 (141)	0.75 (155)	0.76 (53)
Legal structure and property rights (FI)					1.00 (141)	0.68 (135)	0.62 (39)
Property rights (HF)						1.00 (155)	0.54 (43)
Loadings on component 1	0.81	0.81	0.55	0.94	0.85	0.84	–
Loadings on component 2	−0.51	−0.49	0.73	0.16	0.34	−0.04	–

BTI = Bertelsmann Transformation Index; FH = Freedom House; PRS = Political Risk Services; WGI = World Governance Indicators; FI = Fraser Institute; HF = Heritage Foundation; IF = Mo Ibrahim Foundation.
Note: Kendall's tau-b correlation coefficients, N in parentheses.

Table 4. Correlation between the rule of law measures (2005) and measures of neighboring or affiliated concepts

	WGI	PRS	FH	BTI	FI	HF	IF
Electoral self-determination (CIRI)	0.35 (191)	0.08 (137)	0.62 (191)	0.57 (118)	0.29 (139)	0.46 (154)	0.31 (52)
Freedom of speech (CIRI)	0.38 (191)	0.11 (137)	0.61 (191)	0.53 (118)	0.30 (139)	0.44 (154)	0.29 (52)
Freedom of assembly and association (CIRI)	0.40 (191)	0.17 (137)	0.64 (191)	0.51 (118)	0.35 (139)	0.48 (154)	0.31 (52)
Political terror scale (Gibney *et al.*)	−0.58 (176)	−0.49 (137)	−0.54 (176)	−0.45 (119)	−0.52 (138)	−0.55 (153)	−0.55 (53)
Perception of corruption (Transparency International)	0.80 (159)	0.59 (133)	0.57 (157)	0.44 (115)	0.74 (135)	0.69 (146)	0.68 (44)
Level of perceived criminality (Global Peace Index)	−0.50 (121)	−0.43 (114)	−0.34 (120)	−0.25 (90)	−0.54 (109)	−0.37 (115)	−0.47 (26)
Level of violent crime (Global Peace Index)	−0.54 (121)	−0.63 (114)	−0.29 (120)	−0.07 (90)	−0.50 (109)	−0.41 (115)	−0.19 (26)

WGI = World Governance Indicators; PRS = Political Risk Services; FH = Freedom House; BTI = Bertelsmann Transformation Index;
FI = Fraser Institute; HF = Heritage Foundation; IF = Mo Ibrahim Foundation; CIRI = The Cingranelli & Richards Human Rights Dataset.
Note: Kendall's tau-b correlation coefficients, N in parentheses.

empirical convergence of the rule of law indices is relatively low. In a nutshell, different measures of the same concept should have higher correlations than different measures of distinct concepts, even if the same factors cause the phenomena denoted by the distinct concepts (which is quite plausible in the case of democratic elections and the rule of law).

This exercise can also be used in a more constructive way. To unfold the distinction between the identified clusters (FH and BTI vs. FI and HF vs. PRS) in detail, Table 4 also contains a number of other measures affiliated with the rule of law concept: the rights of freedom of speech and freedom of assembly and association as measured by yet other subcomponents from the CIRI, violations of personal integrity rights (such as killings and torture) as measured by the Political Terror Scale provided by Gibney, Cornett, and Wood,[10] the level of criminality and violent crime as measured by two subcomponents of the Global Peace Index,[11] and corruption as measured by the Corruption Perception Index compiled by Transparency International.[12] All data refer to the year 2005, except the two measures of crime from the Global Peace Index that refer to 2007, that is, the first year for which data were available.

The objective of these additional tests is to scrutinize which particular dimensions each of the rule of law indices appears to emphasize. The picture revealed by Table 4 is fairly straightforward. The FH–BTI cluster correlates relatively strongly with the political rights of expression, assembly/association, and free elections but relatively weakly with the measures of crime and corruption. The contrary is the case for the FI–HF cluster. Hence, FH and BTI seem to tap into what could be termed political constitutionalism, whereas FI and HF tap into what is probably best termed public order. Unsurprisingly, the latter tendency is even more pronounced for the PRS, the 'law and order' label of which thus seems apt. Finally, IF and WGI correlate strongly on both dimensions but seem somewhat closer to the order dimension than the constitutionalism dimension, relatively speaking.

This information not only goes to underline that the measures tap different latent dimensions, thereby questioning their interchangeability; it should also assist scholars in choosing the measures best suited for the definition of the rule of law they are concerned with. Insofar as the rule of law is equated with political constitutionalism, FH and BTI seem the superior options. If, *au contraire*, one is on the lookout for a measure of order, PRS probably has the competitive edge, followed by FI and HF, respectively. Finally, should one wish to capture a very encompassing definition of the rule of law – an outlook which clashes with our recommendations in the Conclusions – the composite (and muddled) measures of WGI or IF are preferable.

---

[10] Based on the State Department's Country Reports on Human Rights Practices. The data constructed by Gibney, Cornett, and Wood can be found at http://www.politicalterrorscale.org/.
[11] http://www.visionofhumanity.org/gpi-data/.
[12] http://www.transparency.org/policy_research/surveys_indices/cpi/2005.

Correlations with explanatory variables

The assessment carried out above shows that the rival measures seem to tap into the same empirical phenomenon only to a limited extent. This assertion should of course be qualified by the proposition that this is less so with regard to the pairwise comparison of BTI and FH and those among the FI and HF – and the fact that the WGI seems to be associated with both clusters. Still, the correlations show that the indices are not interchangeable *tout court*.

What is more, the analysis of Casper and Tufis (2003) has demonstrated that even the highly correlated democracy measures may produce different explanatory results when used as dependent variables in cross-national analyses (see also Bollen and Paxton, 2000; Hadenius and Teorell, 2005). One would, *a fortiori*, expect this to be even more pronounced for the rule of law measures.

To carry out this third test of the interchangeability of the indices, we have derived a number of explanatory variables from studies of cross-national differences in compliance with the rule of law (the operationalizations of these variables are presented in Appendix A). These are: oil production (Barro, 2000; Hansson and Olsson, 2006), wealth (Barro, 2000; Joireman, 2004), country size (Hansson and Olsson, 2006), ethno-religious fractionalization (Weingast, 1997; Hayo and Voigt, 2005; Hansson and Olsson, 2006), legal system (Hayek, 1973; Eisenberg, 1988; Joireman, 2004), communist past (Hoff and Stiglitz, 2004; Sandholz and Taagepera, 2005), and religion (Barro, 2000; Hayo and Voigt, 2005).

This exercise does not depend upon including all theoretically relevant independent variables. The objective is not to explain the causes of the rule of law but to test whether the results are relatively similar or dissimilar when the seven measures are used interchangeably as dependent variables in multiple OLS regression analyses.[13] To ease the interpretation, all measures have been calibrated to range from 0 (lowest level of rule of law) to 100 (highest level of rule of law).

In order to investigate whether the results might differ depending on the case coverage, we have run two sets of regressions: first, for all countries included in each of the data sets (Table 5) and, second, for the 116 countries included in five[14] of the seven indices (Table 6). The rule of law data once again covers 2005.

When including all countries, the picture is as follows. A few explanatory variables consistently show a significant (wealth) or non-significant (ethno-religious fractionalization, common law) association with the rule of law. However, the remaining variables exhibit stark dissimilarities depending on the measure used.

The statistically significant associations for some variables (not colonized, communist past, Muslim) even point in opposite directions. Using BTI, we encounter a

---

[13] Owing to lack of variation (which ultimately is due to low country coverage), two of the variables are excluded from the model as regards the IF.

[14] BTI and IF have been excluded as their incorporation would severely limit the number of countries covered by all indices.

Table 5. Summary of regression results with rule of law indices as dependent variable (all countries, 2005)

	WGI	FH	BTI	PRS	FI	HF	IF
Constant	−52.691***	−11.628	−50.585**	−33.558*	−42.185***	−49.142***	3.088
	(11.971)	(19.651)	(21.870)	(20.056)	(10.841)	(16.357)	(35.318)
Oil production	−11.966***	−24.223***	−18.388***	−8.431**	−7.242***	−9.380***	−15.547**
	(2.708)	(4.446)	(4.300)	(3.559)	(2.680)	(2.760)	(7.162)
Wealth	13.484***	10.951***	12.252***	11.133***	11.331***	12.964***	10.274***
	(0.957)	(1.557)	(1.887)	(1.558)	(0.886)	(1.257)	(3.377)
Country size	−0.985**	−1.748**	0.376	−0.492	−0.143	−1.140*	−2.076
	(0.456)	(0.710)	(1.015)	(0.776)	(0.442)	(0.649)	(1.447)
Ethno-religious fractionalization	−3.026	−1.972	2.527	−0.841	0.387	−2.383	3.643
	(4.959)	(8.131)	(10.603)	(6.827)	(5.54)	(−0.390)	(15.142)
Not colonized	9.092***	3.139	−14.220**	11.607***	7.065***	4.981	–
	(2.814)	(4.224)	(6.576)	(4.293)	(1.941)	(3.331)	
Common law	2.283	−2.140	1.542	0.330	2.441	3.843	−2.946
	(2.447)	(3.715)	(5.214)	(3.730)	(2.337)	(2.603)	(15.142)
Communist past	−7.439***	−7.841*	−3.392	8.198**	−0.904	−11.927***	–
	(2.382)	(4.353)	(4.445)	(3.722)	(2.207)	(2.975)	
Muslim	−0.095	−11.464***	−11.000***	9.167**	4.796**	−4.308*	−2.054
	(2.334)	(3.877)	(3.774)	(3.722)	(2.416)	(2.433)	(5.545)
Protestant	5.549*	8.734**	7.739	8.069*	9.033***	9.629***	5.602
	(2.999)	(4.196)	(6.763)	(4.314)	(2.284)	(3.403)	(12.640)
R^2 adjusted	0.71	0.54	0.46	0.50	0.75	0.68	0.34
N	185	180	117	137	138	153	53

WGI = World Governance Indicators; FH = Freedom House; BTI = Bertelsmann Transformation Index; PRS = Political Risk Services;
FI = Fraser Institute; HF = Heritage Foundation; IF = Mo Ibrahim Foundation.
*$P < 0.1$, **$P < 0.05$, ***$P < 0.01$ (two-tailed test). Results refer to regressions with all countries included.
Note: Unstandardized coefficients reported with (heteroscedasticity-consistent) robust standard errors in parentheses.

Table 6. Summary of regression results with rule of law indices as dependent variable (common country coverage, 2005)

	WGI	FH	PRS	FI	HF
Constant	−74.974***	−44.373*	−32.225	−34.858***	−61.442***
	(16.853)	(25.158)	(24.638)	(12.429)	(17.899)
Oil production	−14.285***	−26.552***	−9.856**	−8.221***	−11.513***
	(3.378)	(4.847)	(4.179)	(2.678)	(2.920)
Wealth	15.756***	12.491***	10.914***	10.706***	14.481***
	(1.297)	(1.878)	(2.023)	(0.989)	(1.437)
Country size	−1.038	−0.149	−0.421	−0.254	−1.553**
	(0.650)	(1.056)	(0.857)	(0.522)	(0.704)
Ethno-religious fractionalization	3.867	−0.228	−0.438	−0.431	7.600
	(6.762)	(9.960)	(8.969)	(5.819)	(6.836)
Not colonized	6.628**	−0.642	11.235**	6.850***	4.130
	(3.022)	(5.108)	(5.326)	(2.118)	(3.569)
Common law	−0.728	−5.464	0.459	1.976	2.381
	(2.942)	(4.566)	(4.701)	(2.627)	(3.071)
Communist past	−5.222*	−0.456	6.444*	1.136	−9.951***
	(3.026)	(5.029)	(3.985)	(2.221)	(3.645)
Muslim	4.276	−7.905*	10.632**	5.455**	−3.171
	(2.963)	(4.955)	(4.253)	(2.525)	(2.849)
Protestant	9.912***	12.988***	8.947*	11.635***	10.000***
	(3.294)	(4.741)	(4.575)	(2.216)	(3.636)
R^2 adjusted	0.77	0.58	0.47	0.75	0.71
N	116	116	116	116	116

WGI = World Governance Indicators; FH = Freedom House; PRS = Political Risk Services; FI = Fraser Institute; HF = Heritage Foundation.
*$P < 0.1$, **$P < 0.05$, ***$P < 0.01$ (two-tailed test). Results refer to regressions only including countries covered by all five measures.
Note: Unstandardized coefficients reported with (heteroscedasticity-consistent) robust standard errors in parentheses.

negative relationship with the lack of a colonized past, whereas we find a positive relationship when pinning our faith on WGI, PRS, and FI. Likewise, a communist past produces a significant negative relationship with the rule of law using WGI and HF but a positive relationship using the PRS. Finally, a status as a Muslim country has a strongly significant negative effect on the rule of law using the BTI and FH yet, in contrast, a significant (albeit weaker) positive effect when employing PRS, FI, and HF.

In sum, the association between the explanatory variables and the rule of law ranges from positive to non-existent to negative, depending on which index is used to operationalize the rule of law. Most disturbing are the positive significant associations between PRS and a communist past and a Muslim population, associations which completely contradict the theoretical expectations of the studies from which we have derived these factors.

What happens when we narrow the focus to the 116 countries covered by all the indices (save BTI, which is excluded as it does not cover the OECD countries, and IF, which is excluded as it only covers African countries)? As illustrated in Table 6 below, the results are somewhat more in line with one another. A Communist past and Muslim roots still produce contradictory significant associations, but even this is less pronounced than in Table 5. This indicates that some of the present problems in the literature could be alleviated by confining one's attention to the countries covered by more data sets, that is, by more self-conscious restrictions on the empirical scope of the inquiry.

Non-random patterns of missing data

Such an empirical restriction of the scope of comparison may, however, carry problems of its own. In an assessment of measures of judicial independence, Ríos-Figueroa and Staton (2008) identify systematic patterns of missing data across a selection of indices. Based on their distinct empirical coverage, one might expect that the rule of law measures included in our assessment suffers from the same problem.

To see if this is indeed so, we use the simple test of non-random 'missingness' devised by Ríos-Figueroa and Staton (2008). For each of the independent variables included in the previous regression analyses, we compare the average scores of countries that are included and countries that are missing in four of the data sets, viz. PRS, BTI, FI, and HF. FH and WGI are not assessed because they cover virtually all independent countries of the world, meaning that country data are not missing; IF is excluded because of its focus on one region only.

The question is if there is a statistically significant difference between the scores on each of the independent variables across the two groups (covered vs. missing)? If this is the case, it is likely to introduce biased estimates into any analysis that restricts the scope in accordance with the data availability. The results of the missingness analysis are reported in Table 7.

It turns out that the differences between included and missing countries are indeed significant for a number of the variables. In the cases of wealth, country size, and common law, the differences are significant across the board. For variables such as not colonized, communist past, Muslim, and protestant, the differences are significant for at least one index. Notice also that the direction of the bias – as expressed by the coefficients – sometimes differs. The countries included in BTI are thus significantly poorer than those missing, whereas in the cases of PRS, FI, and HF the countries included are significantly more affluent than their missing equivalents.

Such non-random missingness obviously diminishes our ability to infer from the sample to the general population of all countries.[15] At the very least, scholars

---

[15] Technically, this constitutes truncation on the dependent variable.

Table 7. Simple non-random patterns in missing values in rule of law measures, 2005

	PRS	BTI	FI	HF
Oil production	0.10	0.05	−0.04	0.05
	(0.22–0.13)	(0.22–0.16)	(0.19–0.23)	(0.21–0.16)
Wealth	0.60***	−1.04***	0.52**	0.83***
	(8.77–8.17)	(8.26–9.30)	(8.75–8.23)	(8.73–7.93)
Country size	3.24***	2.81***	2.73***	3.46***
	(12.24–9.00)	(12.40–9.60)	(12.08–9.35)	(12.00–8.54)
Ethno-religious fractionalization	0.01	0.03	0.01	−0.03
	(0.56–0.55)	(0.57–0.54)	(0.56–0.55)	(0.55–0.58)
Not colonized	0.08	−0.14***	0.05	0.03
	(0.15–0.07)	(0.08–0.22)	(0.14–0.09)	(0.14–0.11)
Common law	−0.15**	−0.26***	−0.19**	−0.34***
	(0.29–0.44)	(0.24–0.49)	(0.28–0.47)	(0.27–0.61)
Communist past	0.00	0.29***	0.04	0.19***
	(0.18–0.19)	(0.29–0.00)	(0.19–0.15)	(0.22–0.03)
Muslim	0.06	0.21***	−0.13*	0.07
	(0.28–0.22)	(0.34–0.14)	(0.23–0.36)	(0.28–0.21)
Protestant	−0.10	−0.25***	−0.03	−0.14*
	(0.17–0.28)	(0.11–0.36)	(0.19–0.23)	(0.18–0.32)

PRS = Political Risk Services; BTI = Bertelsmann Transformation Index; FI = Fraser Institute; HF = Heritage Foundation.
$*P < 0.1$, $**P < 0.05$, $***P < 0.01$ (two-tailed test).
Note: Entries are average differences (subtracted means in parentheses) between countries that are covered and missing, respectively, by the rule of law measures.

therefore need to abstain from such inferences or, alternatively, justify why it is possible to infer from the included to the missing countries. Ríos-Figueroa and Staton (2008: 23) note that it is striking that patterns of missingness are almost never dealt with – or even commented upon – by scholars using data sets on judicial independence. Unfortunately, this observation travels well into studies of the rule of law.

The missingness bias afflicts all the data sets reviewed in this article that do not have universal coverage (i.e. all save FH and WGI). Furthermore, and most importantly for our purposes, the dissimilar systematic biases of these indices further undermine their interchangeability.

Conclusion

We have compared seven different indices, which – as may be inferred from the somewhat nonchalant practices present in the literature – have been taken to measure the same phenomenon, viz., the rule of law. As should be clear from our analyses, however, this does not seem to be the case. As soon as we reach the level

of the systematized concept, the indices differ significantly. This is further reflected in the fact that the scores do not correlate highly with each other – at least not vis-à-vis the correlations between the most prominent democracy indices. It is also reflected in the fact that the results of an extensive explanatory model differ starkly, depending on which of the measures is employed as a dependent variable.

As such, the seven indices are clearly not interchangeable. The differences in measurement and coverage of the rule of law indices turn out to have considerable consequences for substantial findings – and much more so than in the neighboring field of democratization studies. The character of the tests carried out in this article means that we are only able to assess the correspondence between the indices in general, and not the extent to which the indicators actually measure the stipulated attributes. Having said that, one particular conclusion concerning the validity of the particular indices does seem warranted. The analyses have shown that PRS is very much out of tune with the other measures, with regard to both conceptual definition and empirical results, indeed, even with regard to its somewhat ambiguous choice of term ('law and order'). Hence, one should be especially cautious of explanatory conclusions about the rule of law based on this index. It is all the more striking that – as shown in the Introduction – this particular measure has until recently been that most often employed in cross-national analyses of the rule of law, due to its wide coverage.

The six other indices also produce relatively dissimilar empirical results when employed in an explanatory model. Yet at least they seem to make up three clusters of indices, each of which captures important aspects of the same phenomenon. First, FH and BTI are so consistently linked that they can almost be used interchangeably in empirical analyses of the rule of law (in the guise of political constitutionalism). Second, FI and HF also show affinities, both concerning correlations and results (including the fact that they seem to tap into order). Third, WGI taps into all the other indices and therefore offers a more general, but also very muddled, measure. This statement also applies to IF.

This could be read as a classical half-full/half-empty conclusion. If PRS is disregarded, the glass seems half full as some interchangeability is allowed for, depending on whether one equates the rule of law with constitutionalism or order or both. However, the problems detected in the analysis of missingness show that the differing empirical scopes of the data sets further undermine their suitability in robustness tests. Most obviously, BTI suffers from massive problems of non-random missingness vis-à-vis FH. Coupled with the differences in conceptualization, correlation, and explanatory results across FH, BTI, WGI, FI, and HF, this makes for a relatively pessimistic conclusion about interchangeability, even within the described clusters.

The general objective of this paper has been to provide easy-to-understand 'knowledge of how the specific measures we select affect the empirical inferences we draw' (Mudde and Schedler, 2010: 413). But what are the more particular analytical consequences of the results? When different data sets of the same

phenomenon produce different results, an explicit validity test is warranted. More particularly, in such a situation we need a basis for selecting certain data sets and not others (cf. Hawken and Munck, 2009). This calls for a careful appraisal of the content validity of the indices vis-à-vis the definition of the rule of law employed by each scholar. However, previous large-N studies exploring the sources of rule of law have frequently neglected this exercise as researchers have tended to select single data sets based on cross-spatial and temporal coverage.

Based on our analysis, it seems fair to expect that the results of such analyses are unlikely to prove robust if other indices are also employed. This means that their results hinge entirely on what the employed rule of law-proxy actually measures and the empirical scope it covers – something that needs careful attention and scrupulous elucidation.

A simple recommendation can be made on this basis. Scholars should be very clear about their definition of the rule of law and they should select data sets in accordance with this definition and with the empirical scope conditions of their theories. As Gerring (1999: 391) pertinently points out, scholars 'have an obligation to state explicitly why (on the basis of which criteria) certain properties and terms were chosen, or excluded'. If this is done, then the lack of consensus may even spur research on the rule of law.

However, if the scholarly community does not in fact wish the rule of law indices to measure different things, then the recommendation changes. In that case, what the ascending research agenda on the rule of law needs now is a definition on the level of the systematized concepts that becomes generally accepted, as did Robert A. Dahl's seminal definition of polyarchy in the democratization literature. Needless to say, this is a necessary but not a sufficient condition, as the subsequent step of measurement also needs to be handled properly.

Until these issues are faced head-on, the problems pestering the rule of law research agenda are likely to subsist. These problems, in turn, speak volumes about the lack of maturity – and cumulativity – of the good governance research field in general (cf. Doornbos, 2003). The very essence of science is, after all, to establish a common language based on sound conceptual premises. Research can only become truly cumulative if the vagueness and ambiguity of ordinary language is reduced. That, in turn, calls for either self-conscious and systematic disagreement about the definition of concepts such as the rule of law or, contrariwise, for explicit agreement. Neither seems to be beckoning at the moment.

Acknowledgements

References

Adcock, R. and D. Collier (2001), 'Measurement validity: a shared standard for qualitative and quantitative research', *American Political Science Review* **95**: 529–546.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat and R. Wacziarg (2003), 'Fractionalization', *Journal of Economic Growth* **8**: 155–194.

Ali, A.M. (2003), 'Institutional differences as sources of growth differences', *Atlantic Economic Journal* **31**: 348–362.

Alvarez, M., J.A. Cheibub, F. Limongi and A. Przeworski (1996), 'Classifying political regime types', *Studies in Comparative International Development* **31**: 3–36.

Andrews, J.T. and G.R. Montinola (2004), 'Veto players and the rule of law in emerging democracies', *Comparative Political Studies* **37**: 55–87.

Arndt, C. and C. Oman (2006), *The Uses and Abuses of Governance Indicators*, OECD: OECD Publishing.

Ashby, N.J. and R.S. Sobel (2008), 'Income inequality and economic freedom in the U.S. States', *Public Choice* **134**: 329–346.

Barahona, D. (2007), 'The freedom house files'. *Monthly Review*, March.

Barro, R. (2000), 'Rule of law, democracy, and economic performance', in O'Driscoll G.P. Jr., Holmes K.R. and Kirkpatrick M. (eds), *2000 Index of Economic Freedom*, Washington and New York: Heritage Foundation and Wall Street Journal, pp. 31–49.

Belton, R.K. (2005), 'Competing definitions of the rule of law', Carnegie Endowment. Carnegie Paper no. 55.

Bertelsmann Transformation Index (2006), 'Bertelsmann Transformation Index'. Retrieved 15 March 2008 from http://www.bertelsmann-transformation-index.de/11.0.html?&L=1

Bleaney, M. and A. Dimico (2008), 'Biogeographical conditions, the transition to agriculture and long-run growth'. CREDIT Research Paper no. 08/15.

Bolaky, B. and C. Freund (2006). 'Trade, regulations, and growth'. Paper presented at the IMF Trade and Growth Conference, January 9, 2006, Washington, DC.

Bollen, K.A. (1992), 'Political rights and political liberties', in T.B. Jabine and R.P. Claude (eds), *Human Rights and Statistics: Getting the Record Straight*, Philadelphia: University of Pennsylvania Press, pp. 188–213.

Bollen, K. and P. Paxton (2000), 'Subjective measures of liberal democracy', *Comparative Political Studies* **31**: 58–86.

Burnside, C. and D. Dollar (2004), 'Aid, policies, and growth'. World Bank Policy Research Working Paper no. 3251.

Butkiewicz, J.L. and H. Yanikkaya (2006), 'Institutional quality and economic growth', *Economic Modelling* **23**: 648–661.

Card, D. and R. Freeman (2002), 'What have two decades of British economic reform delivered?'. Retrieved 7 September 2010 from http://www.nber.org/papers/w8801

Casper, G. and C. Tufis (2003), 'Correlation versus interchangeability: the limited robustness of empirical findings on democracy using highly correlated data sets', *Political Analysis* **11**: 196–203.

Chomsky, N. and E.S. Herman (1988), *Manufacturing Consent*. New York, Scottsville: Pantheon Books.

Coppedge, M. and W. Reinicke (1990), 'Measuring polyarchy', *Studies in Comparative International Development* **25**: 51–72.

Dahl, R.A. (1971), *Polyarchy: Participation and Opposition*, New Haven: Yale University Press.

Doornbos, M. (2003), '"Good governance": the metamorphosis of a policy metaphor', *Journal of International Affairs* **57**: 3–17.

Eisenberg, M.A. (1988), *The Nature of the Common Law*, Cambridge: Harvard University Press.

Fish, M.S. (2002), 'Islam and authoritarianism', *World Politics* **55**: 4–37.

Forewaker, J. and R. Krznaric (2001), 'How to construct a database of liberal democratic performance', *Democratization* **8**: 1–25.

Fraser Institute (2009), 'Economic freedom of the world'. Retrieved 15 November 2009 from http://www.freetheworld.com

Freedom House (2006), 'Freedom in the world'. Retrieved 15 March 2008 from http://www.freedomhouse.org/template.cfm?page=15

Gerring, J. (1999), 'What makes a concept good?', *Polity* 31: 357–393.

Giannone, D. (2010), 'Political and ideological aspects of the measurement of democracy', *Democratization* 17: 68–97.

Goertz, G. (2005), *Social Science Concepts: A Users' Guide*, Princeton: Princeton University Press.

Hadenius, A. and J. Teorell (2005). 'Assessing alternative indices of democracy'. IPSA Concepts & Methods Working Papers 6.

Hansson, G. and O. Olsson (2006), 'Country size and the rule of law'. Working Paper 200, School of Economics: University of Gothenburg.

Hawken, A. and G.L. Munck (2009), 'Do you know your data? Measurement validity in corruption research', Manuscript.

Hayek, F.A. (1973), *Law, Legislation and Liberty*, Chicago: University of Chicago Press.

Hayo, B. and S. Voigt (2005), 'Explaining de facto judicial independence'. Marburg Working Papers on Economics 200507, Philipps-Universität Marburg: Department of Economics.

Heritage Foundation (2009), 'Index of economic freedom'. Retrieved 15 November 2009 from http://www.heritage.org/index

Hoff, K. and J. Stiglitz (2004), 'After the big bang? Obstacles to the emergence of rule of law in post-communist societies', *American Economic Review* 94: 753–763.

IMF (2007), 'Guide on resource transparency'. Retrieved 15 March 2008 from http://www.imf.org/external/np/pp/2007/eng/051507g.pdf

Joireman, S.F. (2001), 'Inherited legal systems and effective rule of law', *Journal of Modern African Studies* 39: 571–596.

—— (2004), 'Colonization and the rule of law', *Constitutional Political Economy* 15: 315–338.

Kaufman, D., A. Kray and M. Mastruzzi (2007), 'Governance matters VI: governance indicators for 1996–2006'. Retrieved 15 March 2008 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=999979

Knack, S. (1996), 'Institutions and the convergence hypothesis', *Public Choice* 87: 207–228.

—— (2002), 'Governance and growth'. The IRIS Discussion Papers on Institutions & Development no. 2/05.

Koelble, T.A. and E. Lipuma (2008), 'Democratizing democracy', *Democratization* 15: 1–28.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999), 'The quality of government', *Journal of Law, Economics, and Organization* 15: 222–279.

Mo Ibrahim Foundation (2010), 'The Ibrahim Index of African Governance'. Retrieved 7 September 2010 from http://www.moibrahimfoundation.org/en/section/the-ibrahim-index

Møller, J. and S.-E. Skaaning (2010), 'Systematizing thin and thick conceptions of the rule of law'. Paper presented at the APSA 2010 Annual Meeting Paper. from SSRN: http://ssrn.com/abstract=1643367

Moore, D. (ed.) (2007), *The World Bank*. KwaZulu-Natal University Press.

Mudde, C. and A. Schedler (2010), 'Introduction: rational data choice', *Political Research Quarterly* 63: 410–416.

Munck, G.L. (2003), 'Measures of democracy, governance and rule of law'. Paper prepared for the World Bank Workshop on Understanding Growth and Freedom from the Bottom Up, July 15–17, Washington, DC.

Munck, G. (2009), *Measuring Democracy*, Baltimore: The Johns Hopkins University Press.

—— (2010), 'Comparative politics: taking stock and looking forward'. Paper prepared for presentation at the 2010 APSA Annual Meeting, September 2–5, Washington, D.C.

Munck, G.L. and J. Verkuilen (2002), 'Conceptualizing and measuring democracy', *Comparative Political Studies* 35: 5–34.

Nanda, P. (2006), 'The "Good Governance" concept revisited', *The ANNALS of the American Academy of Political Science* 603: 269–283.

Norton, S.W. (1998), 'Poverty, property rights, and human well-being', *Cato Journal* 18: 233–245.

Penn World Tables (2007), 'PWT 6.2'. Retrieved 15 March 2008 from http://pwt.econ.upenn.edu/php_site/pwt_index.php

Political Risk Services (2007), 'ICRG'. Retrieved 15 March 2008 from http://www.prsgroup.com/ICRG.aspx

Ríos-Figueroa, J. and J. Staton (2008), 'Unpacking the rule of law'. Committee of Concepts and Methods Working Paper no. 21.

Sandholz, W. and R. Taagepera (2005), 'Corruption, culture, and communism', *International Review of Sociology* **15**: 109–131.

Skaaning, S.-E. (2010), 'Measuring the rule of law', *Political Research Quarterly* **63**: 449–460.

Sunde, U., M. Cervallati and P. Fortunato (2008), 'Are all democracies equally good?', *Economic Letters* **99**(3): 552–556.

Tamanaha, B. (2004), *On the Rule of Law*, Cambridge: Cambridge University Press.

The Economist (2008), *Economics and the Rule of Law*. March 13. http://www.economist.com/node/10849115?story_id=10849115

Thomas, M. (2007), 'What do the worldwide governance indicators measure?'. Retrieved 1 May 2009 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1007527

Uvin, P. (2002), 'On high moral grounds: the incorporation of human rights by the development enterprise', *PRAXIS* **17**: 1–11.

Weingast, B.R. (1997), 'The political foundations of democracy and the rule of law', *American Political Science Review* **91**: 245–263.

Williams, A. and A. Siddique (2008), 'The use (and Abuse) of governance indicators in economics', *Economics of Governance* **9**: 131–175.

World Bank (2008), 'World Development Indicators'. Retrieved 15 March 2008 from http://publications.worldbank.org/online

Appendix A. operationalization of explanatory factors

Oil production

Oil production is operationalized using IMF's (2007) list of hydro-carbon rich countries (2000–05) found in the Guide on Resource Transparency. The countries listed rely heavily on oil production for government revenues and receive the value of 1. The remaining countries receive a 0, meaning that the variable is treated as a dichotomy.

Wealth

Wealth is measured using a standard wealth indicator, namely (natural log of) GDP per capita (purchasing power parity) based on data from the Penn World Tables (2007) for the year 2004.

Country size

The (natural) log of a country's total area in square kilometers is used to measure this variable, based on data from the World Development Indicators provided by the World Bank (2008).

Ethno-religious fractionalization

Data on the degree of heterogeneity are taken from Alesina *et al.*'s (2003) scores on ethnic and religious fractionalization. As the same logic applies to both ethnic

and religious fractionalization, a combined measure has been constructed based on the maximum value of the two original indices for each country.

Not colonized

Data on the colonial past from La Porta *et al*. (1999) have been used to construct a dummy variable distinguishing between countries that have never been colonized and former colonies. The few missing values have been filled in by scores based on information from the CIA's *The World Fact Book*, https://www.cia.gov/library/publications/the-world-factbook.

Common law

The data used to distinguish (predominant) common law countries from countries with another legal system (mainly civil law) are taken from La Porta *et al*. (1999). The few missing values have been filled in by scores based on information from the CIA's *The World Fact Book*, https://www.cia.gov/library/publications/the-world-factbook.

Communist past

A dummy variable for communist and post-communist countries has been constructed.

Dominant religion

Following the procedure used by Steven Fish (2002), we have constructed two dummy variables, distinguishing between countries where Islam and Protestantism, respectively, are the dominant (plurality or majority) religions.