

JRC Statistical Audit of the
WJP Rule of Law Index 2014 

JRC Statistical Audit of the WJP Rule of Law Index[®] 2014

MICHAELA SAISANA AND ANDREA SALTELLI

European Commission Joint Research Centre (Ispra, Italy)

SUMMARY

The JRC analysis suggests that the conceptualized multi-level structure of the WJP Rule of Law Index 2014 is statistically sound in terms of coherence and balance: the overall Index, as well as the eight dimensions, are determined by all underlying components. Furthermore, the analysis has offered statistical justification for the use of equal weights and arithmetic averaging at the various levels of aggregation. Country ranks are also fairly robust to methodological changes related to the estimation of missing data, weighting or aggregation rule (less than ± 3 positions shift with respect to the simulated median in 96% of the cases). The added value of the Rule of Law Index and its dimensions, lays in the ability to summarize different aspects of rule of law in a more efficient and parsimonious manner than would be possible with a collection of almost 500 survey questions taken separately. In fact, the Rule of Law Index, presented this year for the first time as an overall aggregate, has a very high reliability of 0.97 – without being redundant – and captures the single latent phenomenon underlying the eight main dimensions of rule of law.

The WJP Rule of Law Index is intended for a broad audience of policy-makers, civil society, practitioners and academics, and aims at identifying strengths and weaknesses in each country under review and at encouraging policy choices that advance the rule of law. In this respect, the assessment of conceptual and statistical coherence of the Index, and the estimation of the impact of modeling choices on a country's performance are fundamental. They add to the transparency and reliability of the Index, and build confidence in the narratives supported by the measure.

The Econometrics and Applied Statistics Unit at the European Commission Joint Research Centre (JRC) in Ispra, Italy, was invited for a fourth consecutive year by the World Justice Project (WJP) to conduct a thorough statistical assessment of the Index.¹ Fine-tuning suggestions made by the JRC to past releases of the Index were already taken on board by the WJP. The request for a new JRC audit was driven by some re-structuring of the framework, the introduction of the ninth dimension on Informal Justice – measured for the first time this year – and the final aggregation of the eight dimensions into an overall index². The WJP Rule of Law Index was assessed along two main avenues: the statistical coherence of the structure, and the impact of key modeling choices on the Rule of Law Index scores and ranks.

The JRC analysis complements the country rankings for the Rule of Law Index and the underlying dimensions with confidence intervals, in order to better appreciate the robustness of these ranks to the computation methodology. In addition, for the first time this year, the JRC analysis includes an assessment of potential redundancy of information in the Rule of Law framework, and a suggestion on how to monitor changes in the rule of law both in a quantitative and qualitative manner.

1 The JRC analysis was based on the recommendations of the OECD (2008) Handbook on Composite Indicators, and on more recent academic research from the JRC. The JRC auditing studies of composite indicators are available at <http://composite-indicators.jrc.ec.europa.eu/>.

2 The ninth dimension on Informal Justice was presented as part of the conceptual framework for the rule of law but had not been populated with data in past releases of the report. We remind the reader that Informal Justice is not included in the calculation of the overall Index but only used for within country comparisons.

CONCEPTUAL AND STATISTICAL COHERENCE IN THE WJP RULE OF LAW FRAMEWORK

The World Justice Project (WJP), in the fourth release of the 2014 Rule of Law Index, attempts to summarize complex and versatile concepts across 99 countries around the globe with differing social, cultural, economic, and political systems. Modeling the cultural and subjective concepts underlying rule of law at a national scale around the globe raises practical challenges related to the combination of these concepts into a set of numbers and finally into an overall index. Indeed, extending what Saltelli and Funtowisz (2014) argue for models in general, stringent criteria of transparency must be adopted when composite indicators are used as a basis for policy assessments. Failure to open up the black box of composite indicator development is likely to lead only to greater erosion of the credibility and legitimacy of these measures as tools for improved policymaking.

The analysis of conceptual and statistical coherence of an index can be undertaken along four main steps: (a) the consideration of the underlying conceptual framework with respect to the existing literature; (b) the preliminary data quality checks including data coverage, missing values, reporting errors, existence of outliers; (c) the assessment of the statistical coherence through a set of correlation-based analyses, followed by robustness tests about estimation of missing data, weighting schemes and aggregation methods; (d) and finally the qualitative confrontation with the expert bodies in order to get suggestions and reviews about the decisions undertaken in the previous stages of analysis (Saisana, 2011). The WJP team already undertook the first and last steps that are mostly related to the conceptual issues. The JRC audit herein focuses on the second and third steps on the statistical soundness of the Rule of Law Index framework.

DATA CHECKS

The WJP Rule of Law framework builds on nine dimensions, or factors, that are further disaggregated into 47 sub-factors. The scores of these sub-factors are built from almost 500 survey questions drawn from assessments of the general public and local legal experts. Figure 1 illustrates the structure of the 2014 WJP Rule of Law Index.

Country data delivered to the JRC were average scores across experts or individuals along the survey questions (henceforth variables) for 99 countries. These variables are not affected by outliers or skewed distributions³, except for 14 variables spread across six dimensions in the WJP Rule

of Law Index.⁴ Given the high number of variables combined in building a dimension, the skewed distributions of those variables do not bias the results.

A further data quality issue relates to data availability. The 2014 dataset is characterized by excellent data coverage (98% in a matrix of 541 variables × 99 countries). Data availability per dimension and country is also very good or excellent. The WJP, for reasons of transparency and simplicity, calculated sub-factor scores using only available information for each country. This choice, which is common in relevant contexts, might discourage countries from reporting low data values. We tested the implications of 'no imputation' versus the use of the expectation-maximization method for the estimation of missing data and discuss this in the second part of the assessment together with other modeling choices. We anticipate here that some caution is needed in the *Informal Justice*, whereby 24 countries miss values on three or more survey questions (total of eight questions). For most of those countries, the overall score on *Informal Justice* will turn out to be sensitive to the missing data.

PRINCIPAL COMPONENT ANALYSIS AND RELIABILITY ANALYSIS

Principal component analysis (PCA) was used to assess whether the conceptual framework is confirmed by statistical approaches and to identify eventual pitfalls. The analysis confirms the presence of a single statistical dimension in each dimension of the rule of law (one component with eigenvalue greater than 1.0) that captures between 58% (D5: *Order and Security*) up to 88% (D2: *Absence of Corruption*) of the total variance in the underlying sub-factors (Table 1). A more detailed analysis of the correlation structure confirms the expectation that the sub-factors are more correlated to their own dimension than to any other dimension and all correlations are strong and positive. The statistical reliability, measured by the Cronbach-alpha (or c-alpha), is very high at 0.90 (up to 0.95) for seven of the nine dimensions, which is well above the 0.7 threshold for a reliable aggregate (see Nunnally, 1978). Instead, reliabilities are low for *Order and Security* (c-alpha = .62), and *Informal Justice* (c-alpha = .36). Nevertheless, once excluding a specific sub-factor from either dimension (#5.2: *civil conflict is effectively limited* from *Order and Security* and #9.1: *informal justice is timely and effective* from *Informal Justice*), the reliabilities of the two dimensions enter within the recommended limits (0.70 or slightly above, see Table 1).

Overall, the conceptual grouping of sub-factors into dimensions is statistically supported by the data for seven

³ Groeneveld and Meeden (1984) set the criteria for absolute skewness above 1 and kurtosis above 3.5. The skewness criterion was relaxed to 'above 2' to account for the small sample (99 countries).

⁴ In the WJP Rule of Law Index 'sub-factors' are equivalent to sub-dimensions.

FIGURE 1. SCHEMATIC REPRESENTATION OF THE 2014 RULE OF LAW FRAMEWORK AND INDEX.



Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.
Notes: Rearranged from the information provided on the WJP Rule of Law Index 2014 main report.

dimensions of the rule of law, whilst a careful revision is needed for *Order and Security* and *Informal Justice*.

Furthermore, the analysis suggests that the eight dimensions (D1 to D8) share a single latent factor that captures 83% of the total variance and their aggregate has a reliability of 0.97. Instead, the *Informal Justice* (D9) is almost orthogonal

(not related) either to any of the eight dimensions or to the overall index. The revision suggested above for this dimension (*i.e.* to exclude #9.1: *informal justice is timely and effective*), whilst sufficient to render the dimension a reliable average, it does not suffice to render it coherent to the other eight dimensions.

TABLE1: STATISTICAL COHERENCE IN THE 2014 RULE OF LAW INDEX

RULE OF LAW DIMENSIONS	VARIANCE EXPLAINED	C-ALPHA	C-ALPHA WHEN EXCLUDING ONE COMPONENT							
			#.1	#.2	#.3	#.4	#.5	#.6	#.7	#.8
Rule of Law Index	83	.97	.96	.96	.96	.97	.97	.96	.96	.96
1: Constraints on Government Powers	83	.95	.94	.93	.94	.95	.94	.94		
2: Absence of Corruption	88	.96	.92	.94	.93	.95				
3: Open Government	78	.89	.89	.84	.87	.87				
4: Fundamental Rights	73	.95	.94	.93	.93	.93	.94	.93	.94	.94
5: Order and Security	58	.62	.30	.73	.44					
6: Regulatory Enforcement	79	.93	.91	.90	.92	.91	.92			
7: Civil Justice	66	.91	.90	.89	.87	.88	.91	.88	.89	
8: Criminal Justice	77	.95	.94	.93	.93	.94	.93	.95	.93	
9: Informal Justice	69	.37	.69	.00	.04					

Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.
Notes: (1) Column "Variance explained" shows the amount of total variance explained by the first principal component across the eight dimensions of the Index, or the sub-factors in the case of the dimensions. (2) c-alpha or Cronbach-alpha is a measure of statistical reliability (values greater than 0.7 are recommended for good reliability). (3) Informal Justice is not included in the calculation of the Rule of Law Index but only in the framework of rule of law.

Concluding, the results from this analysis could be used as a statistical justification for the WJP choice to aggregate further the eight dimensions into a single index by using an arithmetic average, and not to include *Informal Justice* in the index calculation, but to use it instead only for within country comparisons.

WEIGHTS AND IMPORTANCE

Next, tests focused on identifying whether the Rule of Law dimensions and the overall Index are statistically well-balanced in the underlying components. In the present context given that all dimensions are built as simple arithmetic averages (*i.e.* equal weights for the relative sub-factors), and the index as a simple average of the eight dimensions, our analysis answers the question: ‘are the sub-factors – or the dimensions – really equally important?’ We used an importance measure (henceforth S_i), most known as correlation ratio, which is the non-linear equivalent to the classical Pearson correlation coefficient (Saltelli *et al.*, 2008). The S_i describes ‘the expected reduction in the variance of the eight dimension scores that would be obtained if a given sub-factor could be fixed’. As discussed in Paruolo *et al.*, 2013, we can take this as a measure of importance⁵; thus if sub-factors are supposed to be equally important their S_i values should not differ too much. Results are reassuring: all sub-factors are important in classifying countries within each dimension, though some sub-factors are slightly more important than others (Table 2). Although still acceptable, the least coherent results are: under *Fundamental Rights* dimension, the contribution of the sub-factor 4.1 (*equal treatment and absence of discrimination*) and 4.5 (*freedom of belief and religion is effectively guaranteed*) compared to the remaining sub-factors on the basis of the lower importance. Similarly, sub-factors 5.2 (*civil conflict is effectively limited*), sub-factor 7.5 (*civil justice is not subject to unreasonable delays*) and sub-factor 9.1 (*informal justice is timely and effective*) have a lower contribution to the variance of the respective dimension compared to the other underlying sub-factors. Finally, all eight dimensions are roughly equally important in determining the variation in the Index scores, though *Order and Security* (D5) is slightly less influential. All together the degree of coherence of the Index is remarkable, *i.e.* all dimensions and the overall index appear to be balanced and coherent.

⁵ The Pearson correlation ratio or first order sensitivity measure offers a precise definition of importance, that is ‘the expected reduction in variance of the CI that would be obtained if a variable could be fixed’; it can be used regardless of the degree of correlation between variables; it is model-free, in that it can be applied also in non-linear aggregations; it is not invasive, in that no changes are made to the index or to the correlation structure of the indicators.

ASSESSING POTENTIAL REDUNDANCY OF INFORMATION IN THE RULE OF LAW DIMENSIONS

A very high statistical reliability may be the result of redundancy of information in an aggregate. This is not the case in the Rule of Law Index. The high statistical reliability ($c\text{-alpha} = 0.97$) of the simple average of the eight dimensions is a sign of a sound composite indicator that brings additional information on the rule of law issues in the countries around the world. This is shown in Table 3, which presents, for all pairwise comparisons between the Index and the dimensions, the Spearman rank correlation coefficients (above the diagonal) and the percentage of countries that shift 10 positions or more (below the diagonal). In fact, of the 99 countries included this year, for almost 30% (up to 53%) of the countries, the Index ranking and any of the eight dimension rankings differ by 10 positions or more. This is a desired outcome because it evidences the added value of the Index ranking as a benchmarking tool, namely to help highlighting aspects of rule of law that do not emerge directly by looking into the eight dimensions separately.

IMPACT OF MODELING ASSUMPTIONS ON THE WJP RULE OF LAW INDEX RESULTS

The WJP Rule of Law Index and the underlying dimensions are the outcome of choices: the framework (driven by theoretical models and expert opinion), the variables included, the estimation or not of missing values, the normalization of the variables, the weights assigned to the variables and sub-factors, and the aggregation method, among other elements. Some of these choices are based on expert opinion, or common practice, driven by statistical analysis or the need for ease of communication. The aim of the uncertainty analysis is to assess to what extent – and for which countries in particular – these choices might affect country classification. We have dealt with these uncertainties simultaneously in order to assess their joint influence and fully acknowledge their implications (Saltelli and D’Hombres, 2010). Data are considered to be error-free since the WJP team already undertook a double-check control of potential outliers and eventual errors and typos were corrected during this phase.

The robustness assessment of the WJP Rule of Law Index was based on a combination of a Monte Carlo experiment and a multi-modeling approach. This type of assessment aims to respond to eventual criticism that the country scores associated with aggregate measures are generally not calculated under conditions of certainty, even if they are frequently presented as such (Saisana *et al.*, 2005, 2011). The Monte Carlo simulation related to the weights and comprised 1,000 runs, each corresponding to a different set of weights of the sub-factors underlying each dimension, randomly

TABLE 2: IMPORTANCE MEASURES (VARIANCE-BASED) FOR THE SUB-FACTORS AND DIMENSIONS IN THE 2014 WJP RULE OF LAW INDEX.

	#.1	#.2	#.3	#.4	#.5	#.6	#.7	#.8
INDEX	0.87 [.84, .91]	0.93 [.92, .95]	0.87 [.86, .9]	0.8 [.76, .86]	.63* [.54, .67]	0.95 [.94, .96]	0.87 [.87, .92]	0.88 [.87, .9]
D1	0.91 [.88, .92]	0.78 [.77, .82]	0.71 [.7, .75]	0.82 [.75, .85]	0.88 [.82, .89]			
D2	0.95 [.93, .96]	0.87 [.86, .91]	0.95 [.9, .95]	0.8 [.8, .86]				
D3	0.7 [.69, .78]	0.87 [.84, .9]	0.76 [.75, .83]	0.83 [.82, .87]				
D4	.57* [.56, .6]	0.9 [.85, .9]	0.74 [.73, .79]	0.79 [.74, .85]	.61* [.56, .65]	0.88 [.83, .9]	0.81 [.7, .84]	0.75 [.74, .79]
D5	0.66 [.66, .76]	.38* [.38, .44]	0.66 [.63, .72]					
D6	0.83 [.81, .84]	0.88 [.85, .9]	0.72 [.72, .8]	0.81 [.8, .86]	0.75 [.69, .81]			
D7	0.59 [.59, .62]	0.67 [.63, .73]	0.82 [.79, .84]	0.76 [.73, .83]	.39* [.39, .5]	0.77 [.77, .83]	0.67 [.66, .72]	
D8	0.65 [.64, .71]	0.8 [.77, .87]	0.8 [.79, .87]	0.7 [.7, .73]	0.89 [.86, .91]	0.76 [.69, .85]	0.84 [.83, .88]	
D9	.43* [.42, .6]	0.7 [.7, .79]	0.66 [.66, .81]					

Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.

Notes: (1) Numbers represent the kernel estimates of the Pearson correlation ratio, as in Paruolo et al., 2013. Min-max estimates for the Pearson correlation ratio derive from the choice of the smoothing parameter and are shown in parenthesis. (2) Sub-factors that have much lower contribution to the variance of the relevant Dimension scores than the equal weighting expectation are marked with an asterisk. (3) D1: Constraints on Government Powers, D2: Absence of Corruption, D3: Open Government, D4: Fundamental Rights, D5: Order and Security, D6: Regulatory Enforcement, D7: Civil Justice, D8: Criminal Justice, D9: Informal Justice.

sampled from uniform continuous distributions centered in the reference values. The choice of the range for the weights' variation was driven by two opposite needs: on the one hand, the need to ensure a wide enough interval to have meaningful robustness checks (about $\pm 25\%$ of the reference value); on the other hand, the need to respect the rationale of the WJP that the sub-factors have roughly the same importance when calculating a dimension. Given these considerations, limit values of uncertainty intervals have been defined as shown in Table 4.

The multi-modeling approach involved combinations of the remaining two key assumptions on the 'no imputation' of missing data and the aggregation formula across the sub-factors or the dimensions. The WJP calculated sub-factor scores using only available information for each country⁶. This choice (often termed as 'no imputation') was confronted with the application of the expectation-maximization method

⁶ Note that here 'no imputation' is equivalent to replacing missing values with the average of the available data within each sub-factor.

for the estimation of the missing data⁷. Regarding the WJP assumption on the aggregation function (arithmetic average), and despite the fact that it received statistical support (see principal component analysis results in the previous section), decision-theory practitioners have challenged this type of aggregation because of their fully compensatory nature, in which a comparative advantage of a few variables can compensate a comparative disadvantage of many variables (Munda, 2008). This offsetting might not be always desirable when dealing with fundamental aspects of rule of law. Hence, we considered the geometric average instead, which is a partially compensatory approach.⁸ Consequently, we tested

⁷ The Expectation-Maximization (EM) algorithm (Little and Rubin, 2002) is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating two steps: (1) The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates. (2) The maximization M-step: Given a complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step. The two steps are iterated until the iterations converge.

⁸ In the geometric average, sub-factors are multiplied as opposed to summed in the arithmetic average. Sub-factor weights appear as exponents in the multiplication. To avoid that zero values introduce a bias in the geometric average, we re-scaled linearly the sub-factors scores to a minimum of 0.01.

TABLE 3: ADDED-VALUE OF THE RULE OF LAW INDEX.

	INDEX	D1	D2	D3	D4	D5	D6	D7	D8	D9
INDEX		0.88	0.92	0.89	0.86	0.77	0.94	0.89	0.91	0.16
D1	42		0.75	0.85	0.86	0.52	0.82	0.75	0.76	0.18
D2	31	54		0.79	0.72	0.72	0.91	0.84	0.89	0.18
D3	36	46	52		0.83	0.63	0.84	0.74	0.75	0.14
D4	38	48	54	52		0.53	0.78	0.74	0.72	0.07
D5	53	64	56	58	67		0.71	0.71	0.77	0.05
D6	30	40	28	40	54	60		0.89	0.83	0.17
D7	31	47	44	54	52	61	35		0.83	0.21
D8	35	54	40	56	58	54	38	44		0.18
D9	78	75	77	73	77	82	72	76	75	

Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.

Notes: (1) Numbers above the diagonal are Spearman rank correlation coefficients, whilst numbers below the diagonal are the percentage of countries that shift +10 positions between the rankings. (3) D1: Constraints on Government Powers, D2: Absence of Corruption, D3: Open Government, D4: Fundamental Rights, D5: Order and Security, D6: Regulatory Enforcement, D7: Civil Justice, D8: Criminal Justice, D9: Informal Justice.

four models based on the combination of no imputation versus expectation-maximization and arithmetic versus geometric average. Combined with the 1,000 simulations per model to account for the uncertainty in the weights across the sub-factors, we carried out altogether 4,000 simulations.

Selected results of the uncertainty analysis are provided in Figure 2, which shows median ranks and 90% intervals computed across the 4,000 Monte Carlo simulations for the overall Index and for two dimensions: Absence of Corruption (D2, one of the most robust dimensions) and Order and Security (D5, one of the least robust dimensions). Countries are ordered from the highest to the lowest levels of rule

of law according to their reference rank in the WJP (black line), the dot being the simulated median rank. Error bars represent, for each country, the 90% interval across all simulations.

More specifically, if one takes the simulated median rank as being representative of these scenarios, then the fact that the dimension ranks are close to the median ranks suggests that the eight dimensions and the overall Index are suitable summary measures of the rule of law aspects. Country ranks in the overall Index and in all eight dimensions are very close to the median rank: 90 percent of the countries shift with respect to the simulated median less than ± 1 position in

TABLE 4: UNCERTAINTY PARAMETERS (MISSING VALUES, WEIGHTS AND AGGREGATION FUNCTION)

	REFERENCE	ALTERNATIVE
I. UNCERTAINTY RELATED TO MISSING DATA	NO ESTIMATION OF MISSING DATA	EXPECTATION MAXIMIZATION (EM)
II. UNCERTAINTY IN THE AGGREGATION FUNCTION	ARITHMETIC AVERAGE	GEOMETRIC AVERAGE
	REFERENCE VALUE FOR THE WEIGHT	DISTRIBUTION FOR UNCERTAINTY ANALYSIS
III. UNCERTAINTY INTERVALS FOR THE EIGHT DIMENSION WEIGHTS	0.125	U[0.094, 0.156]
IV. UNCERTAINTY INTERVALS FOR THE SUB-FACTOR WEIGHTS		
1: CONSTRAINTS ON GOVERNMENT POWERS (6 SUB-FACTORS)	0.167	U[0.125, 0.208]
2: ABSENCE OF CORRUPTION (4 SUB-FACTORS)	0.250	U[0.188, 0.313]
3: OPEN GOVERNMENT (4 SUB-FACTORS)	0.250	U[0.188, 0.313]
4: FUNDAMENTAL RIGHTS (8 SUB-FACTORS)	0.125	U[0.094, 0.156]
5: ORDER AND SECURITY (3 SUB-FACTORS)	0.333	U[0.250, 0.417]
6: REGULATORY ENFORCEMENT (5 SUB-FACTORS)	0.200	U[0.150, 0.250]
7: CIVIL JUSTICE (7 SUB-FACTORS)	0.143	U[0.107, 0.179]
8: CRIMINAL JUSTICE (7 SUB-FACTORS)	0.143	U[0.107, 0.179]

Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.

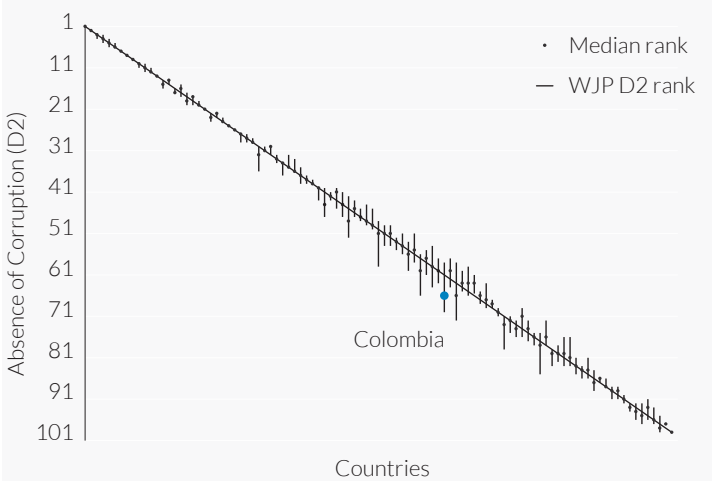
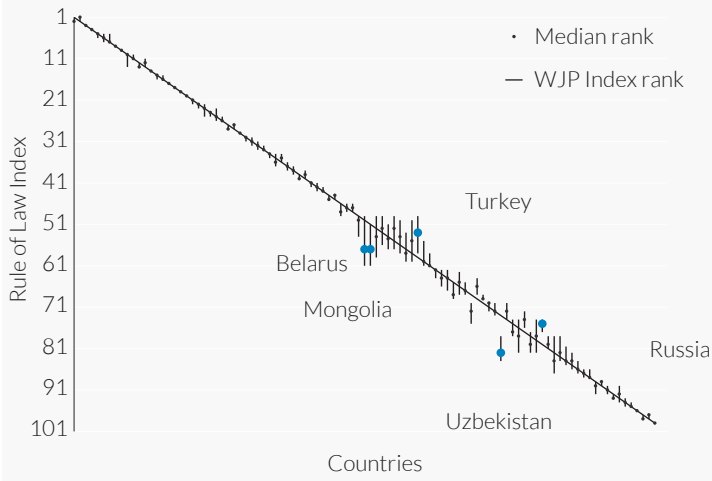
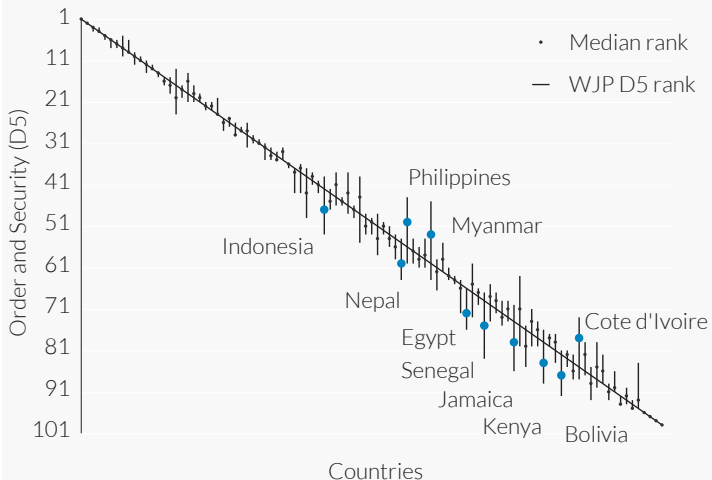
the Rule of Law Index, Constraints on Government Powers (D1) and Fundamental Rights, (D4); less than ± 2 positions in Absence of Corruption (D2), Open Government (D3), Regulatory Enforcement (D6) and Criminal Justice (D8); less than ± 3 positions in Civil Justice (D7); and less than ± 5 positions in Order and Security (D5). These moderate shifts for the vast majority of the countries can be taken as an indication that country classifications along the rules of law issues depend mostly on the variables used and not on the methodological judgments made during the aggregation.

Simulated intervals for most countries are narrow enough, hence robust to changes in the estimation of missing data, weights and aggregation formula – less than 6 positions in 75% of the cases across the eight dimensions and the overall Index. These results suggest that for the vast majority of the countries, the Rule of Law Index ranks allow for meaningful inferences to be drawn.

Nevertheless, few countries have relatively wide intervals (more than 15 positions): none on Constraints on Government Powers (D1), Absence of Corruption (D2), Fundamental Rights (D4), Civil Justice (D7); China, Malaysia, and United Arab Emirates on Open Government (D3); Cote d'Ivoire, Jamaica, Myanmar, Philippines, Russia, Senegal, and Thailand on Order and Security (D5); Uzbekistan on Regulatory Enforcement (D6); and Brazil and Panama on Criminal Justice (D8). These relatively wide intervals are due to compensation of low performance on some sub-factors with a very good performance on other sub-factors in a given dimension (see country profiles in the main part of the report). These cases have been flagged herein as part of the uncertainty analysis in order to give more transparency in the entire process and to help appreciate the WJP Rule of Law Index results with respect to the choices made during the development phase. To this end, Table 5 reports the Index and dimension ranks together with the simulated intervals (90% of the 4000 scenarios capturing estimation of missing data, weights and aggregation formula).

The fact that the dimension on Absence of Corruption (D2) is one of the most robust dimensions in the WJP Rule of Law Index with respect to modeling assumptions and also very coherent – as discussed in the previous section, see Table 1 and Table 2 – is all the more noteworthy given its inclusion in the Corruption Perception Index of Transparency International, as one of the thirteen measures describing perception of corruption in the public sector and among politicians.

FIGURE 2: UNCERTAINTY ANALYSIS (WJP INDEX AND SELECTED DIMENSION RANKS VS. MEDIAN RANK, 90% INTERVALS).



Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.
Notes: Countries are ordered from high to low levels of rule of law. Median ranks and intervals are calculated over 4,000 simulated scenarios combining random weights (25% above/below the equal weights assumption), imputed versus missing values, and geometric versus arithmetic average at the dimension (or sub-factor) level. Countries with less reliable ranks – 5 or more positions shift from the simulated median rank are flagged.

Though country rankings are not calculated by the WJP for the Informal Justice, a similar robustness analysis reveals that twenty two countries in this dimension have relatively wide intervals (more than 15 positions)⁹. These wide intervals are in most cases due to the amount of missing data (4 or more out of the 8 question items). This outcome further supports the WJP choice to use the Informal Justice dimension scores as an indication for within country comparisons and not across countries.

As a general remark, the robustness of an index should not be interpreted as an indication of the index's quality. It is instead a consequence of the index's dimensionality. In other words, robustness is to some extent the flip side of redundancy: a very high correlation between variables will lead to an index ranking that is practically not affected by the methodological choices, so the index will be both robust and redundant. Similarly, a low correlation among variables would imply that the methodological choices are very important in determining country rankings, and thus the index is unlikely to be robust to these choices. The results herein have revealed that the 2014 Rule of Law Index is robust without being redundant.

RULE OF LAW INDEX AND THE VARIABILITY OF ITS DIMENSIONS

Finally, we study the relationship between the Rule of Law Index scores of a given country and the variability of its eight underlying dimensions, namely what the relationship is, if any, between the Index score and a balanced performance in constraints on government powers, absence of corruption, open government, fundamental rights, order and security, regulatory enforcement, civil justice, and criminal justice. While the Index values provide a quantitative indication of trends in rule of law, changes in the dimension's variability convey information on the quality of the changes: an increase in rule of law may be achieved by improving the performance in specific dimensions, whilst a decrease in the coefficient of variation may be achieved by reducing gaps in performance between dimensions.

As can be seen from the scissor's pattern in Figure 3, generally countries with higher levels of rule of law exhibit less variability since they tend to achieve high values in most of the underlying dimensions. The opposite generally holds true for countries with lower levels of rule of law. The average variability in the top tertile group is 0.11, in the middle tertile group is 0.21, and in the low tertile group is 0.27. This reflects the fact that countries with lower levels of rule of law generally display larger discrepancies in performance

between dimensions, and that focusing only in particular dimensions while allowing performance gaps between dimension yields only marginal results in their overall rule of law score. However, it is worth noting that there is a certain variance in the results: although Tanzania and Pakistan belong to the low tertile group in the rule of law, their variability is just above the average variability of the top tertile group. The same applies to a number of countries in the middle tertile group (South Africa, Colombia, and Macedonia-FYR). Instead, although the United Arab Emirates belongs to the top tertile group, its variability is above the average of the middle tertile group.

The Pearson correlation coefficient between the Rule of Law Index and the coefficient of variation is -0.83, what reflects a high degree of negative association between the Index and the variability of its eight dimensions.

CONCLUSIONS

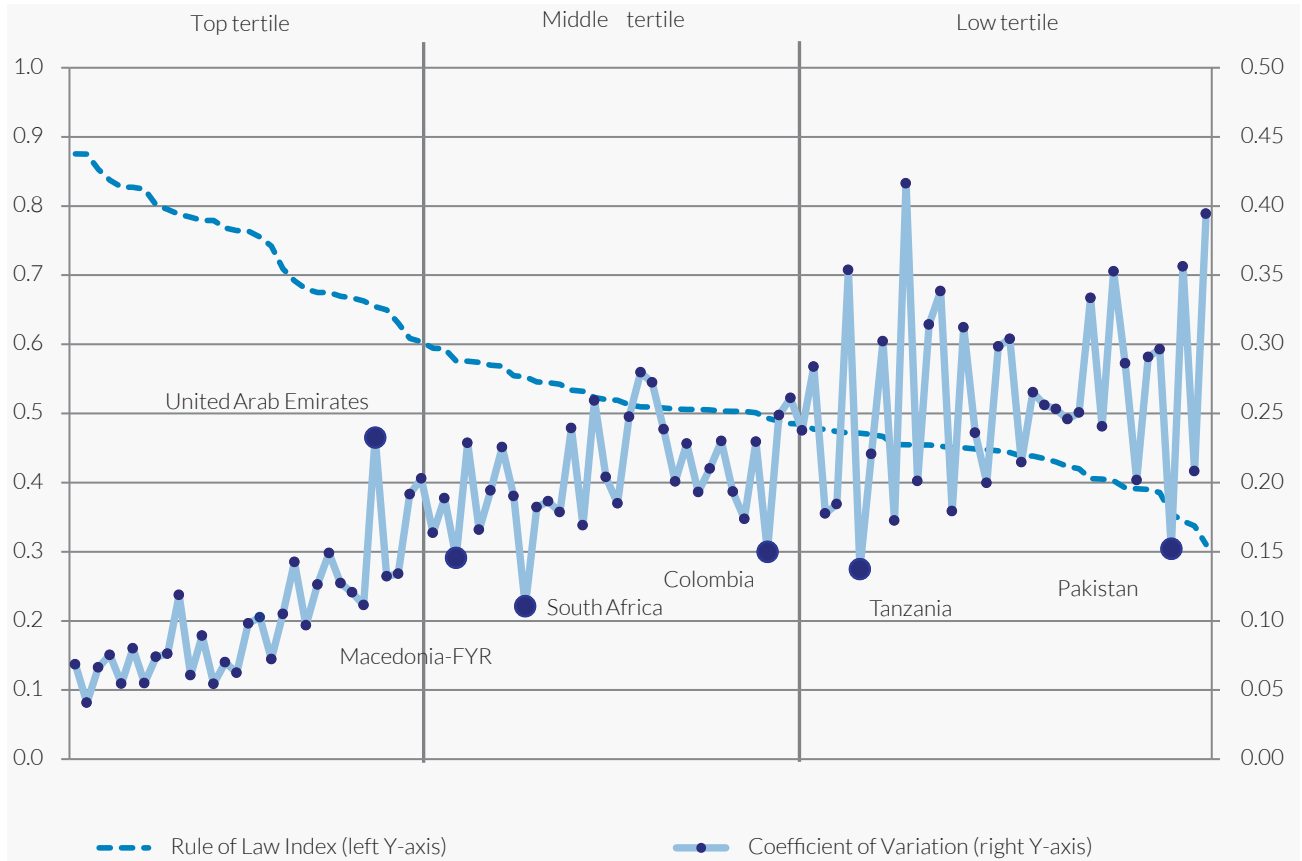
The WJP team invited the JRC for the fourth consecutive year to delve into the statistical properties of the revised Rule of Law Index, so as to ensure the transparency and reliability of the results and to enable academics and policymakers to derive more accurate and meaningful conclusions. In fact, stringent criteria of transparency must be adopted when composite indicators are used as a basis for policy assessments. Failure to open up the black box of composite indicator development is likely to lead only to greater erosion of the credibility and legitimacy of these measures as tools for improved policymaking.

The JRC analysis suggests that the conceptualized multi-level structure of the 2014 WJP Rule of Law Index – calculated through almost 500 survey questions and eight dimensions for 99 countries – is statistically sound, coherent, and balanced. Indeed, within each dimension a single latent factor is identified and all sub-factors are roughly equally important in determining the variation of the respective dimension scores. This outcome can be used as statistical justification for the equal weights and arithmetic averaging at the various levels of aggregation of the Rule of Law Index – which should not be taken for granted when arithmetic averaging is concerned. The Absence of Corruption dimension is especially coherent and robust, which is noteworthy given its inclusion in the Corruption Perception Index of Transparency International.

Country ranks across the eight dimensions and in the overall Index are also fairly robust to methodological changes related to the estimation of missing data, weighting or aggregation rule (less than ± 3 positions shift in 96% of the cases). Consequently, benchmarking inferences can be drawn

⁹ These are: Albania, Australia, Chile, Croatia, Czech Republic, Finland, Greece, Hong Kong SAR of China, Hungary, Iran, Italy, Jamaica, Macedonia-FYR, New Zealand, Norway, Portugal, Singapore, Slovenia, Sri Lanka, Turkey, United Kingdom, and Uruguay.

FIGURE 3: RULE OF LAW INDEX VALUES AND THE VARIABILITY OF THEIR UNDERLYING DIMENSIONS.



Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law Index 2014.
Notes: Countries are ordered from high to low levels of rule of law. The coefficient of variation for each country is calculated as the ratio of the standard deviation across the eight dimensions of the rule of law to their average.

for most countries in the Rule of Law Index and the eight underlying dimensions, whilst some caution may be needed for a few countries. Note that perfect robustness would have been undesirable as this would have implied that the Index and the dimensions are perfectly correlated and hence redundant, which is not the case. In fact, one way in which the 2014 Rule of Law Index helps to highlight other aspects of rule law is by pinpointing the differences in rankings that emerge from a comparison between the Index and each of the eight dimensions: for more than 30% (up to 53%) of the countries, the Index ranking and any of the eight dimensions rankings differ by 10 positions or more.

The main refinements suggested by the present analysis relate to the dimensions of Order and Security and Informal Justice. The former needs a revision with respect to the sub-factor on civil conflict is effectively limited, whilst Informal Justice appears to be measuring an aspect of the rule of law that is totally different to what is being measured by the other eight dimensions. The missing data for 20+ countries within Informal Justice do not allow for a reliable estimation of their performance level in this respect. Beyond conceptual issues, these statistical considerations may justify the WJP’s

choice not to include Informal Justice in the index calculation, but to consider it instead indicatively for within country comparisons only.

The added value of the 2014 WJP Rule of Law Index and its underlying dimensions — developed using international quality standards and tested using state of the art statistical analyses — lays in the ability to summarize different aspects of rule of law in a more efficient and parsimonious manner than what is possible with a collection of almost 500 survey questions taken separately. In fact, the Rule of Law Index, presented this year for the first time as an overall aggregate, has a very high reliability 0.97 and captures indeed the single latent phenomenon underlying the eight main dimensions of rule of law. In past reports, the WJP team had opted not to calculate an overall index in order to shed more light onto the dimensions of the rule of law. Hopefully, this year’s initiative to accompany the detailed country profiles with an overall rule of law score will reinforce the media’s uptake of the Rule of Law Index and the WJP’s engagements with civil society.

REFERENCES

Groeneveld, R. A., Meeden, G. 1984. Measuring skewness and kurtosis. *The Statistician* 33: 391–99.

Little, R. J. A., Rubin, D. B. 2002. *Statistical Analysis with Missing Data*. 2nd edition. Hoboken, NJ: John Wiley & Sons.

Munda, G. 2008. *Social Multi-Criteria Evaluation for a Sustainable Economy*. Berlin Heidelberg: Springer-Verlag.

Nunnally, J. 1978. *Psychometric theory*. New York: McGraw-Hill.

OECD/EC JRC, 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD.

Paruolo P., Saisana M., Saltelli A. 2013, Ratings and Rankings: voodoo or science? *J Royal Statistical Society A* 176(3): 609-634.

Saisana, M., D'Hombres, B., Saltelli, A. 2011. Ricketty Numbers: Volatility of University Rankings and Policy Implications. *Research Policy* 40: 165–77.

Saisana, M., and Saltelli, A. 2011. Rankings and Ratings: Instructions for use. *Hague Journal on the Rule of Law* 3(2): 247-268.

Saisana, M. 2011. Statistical tests on the Global Innovation Index (p.57-64), in Dutta, S. (Ed), *The Global Innovation Index 2011. Accelerating Growth and Development*, INSEAD.

Saisana, M., Saltelli, A., Tarantola, S. 2005. Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators. *Journal of the Royal Statistical Society A* 168 (2):307–323.

Saltelli, A., Funtowicz S. 2014. When All Models Are Wrong, Computer Modeling, *Issues in Science and Technology*, 79-85.

Saltelli, A., D'Hombres, B. 2010. Sensitivity Analysis Didn't Help. A Practitioner's Critique of the Stern Review. *Global Environmental Change* 20: 298–302.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. 2008. *Global Sensitivity Analysis: The Primer*. Chichester, England: John Wiley & Sons.