



# **Part III: Statistical Tests on the WJP Rule of Law Index**

# Statistical Tests on the WJP Rule of Law Index

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 is statistically coherent and no dimension is dominated by any of its underlying components. Country ranks across the eight dimensions are also fairly robust to methodological changes related to the estimation of missing data, weighting or aggregation rule (less than  $\pm 1$  position shift in 90% of the cases).

The assessment of conceptual and statistical coherence of the World Justice Project (WJP) Rule of Law Index and the estimation of the impact of modeling assumptions on a country's performance are useful steps: they add to the transparency and reliability of the Index and build confidence in the narratives supported by the measure. Modeling the cultural and subjective concepts underlying the rule of law at a national scale around the globe raises practical challenges related to the combination of these concepts into a single set of numbers.

The Econometrics and Applied Statistics Unit at the European Commission Joint Research Centre in Ispra, Italy has undertaken for a second consecutive year, and upon request of the WJP, a thorough statistical assessment of the Index.<sup>1</sup> The WJP Rule of Law Index was assessed along two main avenues: the conceptual and statistical coherence of its structure, and the impact of key modeling assumptions on its scores and ranks.

## Conceptual and statistical coherence in the WJP Rule of Law framework

Country data delivered to the JRC represented average scores of public or expert opinion on 479 variables. These variables are not affected by outliers or skewed distributions<sup>2</sup>, except for 16 variables spread across six factors in the WJP Rule of Law Index.<sup>3</sup> Given the high number of variables combined in building a factor, the skewed distributions of those variables do not bias the results. Some reservations on Civil conflict is effectively limited (sub-factor 3.2) are discussed later. The 2011 dataset is characterized by excellent data coverage (92% in a matrix of 479 variables × 66 countries). Data coverage per dimension and country is also very good or excellent. A further data quality issue relates to the treatment of missing values. The WJP, for reasons of transparency and replicability, 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 hot-deck imputation method and discuss this in the second part of the assessment together with other modeling assumptions.

Principal component analysis (PCA) was used to assess to what extent the conceptual framework is

confirmed by statistical approaches and to identify eventual pitfalls. The analysis confirms the WJP Rule of Law Index structure, as within each of the eight dimensions the first latent factor captures between 55% up to 93% of the variance (best result for *Absence of Corruption* – Factor 2). 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. Hence, no-reallocation of sub-factors is needed. Finally, the eight factors share a single latent factor that captures 82% of the total variance. This latter result could be used as a statistical justification for aggregating further the eight dimensions into a single index by using a weighted arithmetic average. This is not currently done, as the WJP team aims to shed more light on the dimensions of the rule of law as opposed to an overall index.

Next, tests focused on identifying whether the eight dimensions of the WJP Rule of Law Index are statistically well-balanced in the underlying sub-factors. In the present context given that all dimensions are built as simple arithmetic averages (*i.e.* equal weights for the relative sub-factors), our analysis answers the question: ‘*are the sub-factors really equally important?*’ We used an ‘importance measure’ (henceforth  $S_i$ ), known as correlation ratio or first order sensitivity measure (Saltelli *et al.*, 2008). The  $S_i$  describes ‘the expected reduction in the variance of factor scores that would be

**Table 1. Importance measures (variance-based) for the eight factors of the WJP Rule of Law Index**

Sub-factor	F1	F2	F3	F4	F5	F6	F7	F8
#.1		0.88 (0.05)	0.80 (0.08)	0.6 (0.11)	0.64 (0.1)	0.87 (0.07)		0.64 (0.10)
#.2	0.87 (0.08)	0.93 (0.06)	0.27 (0.18)*	0.64 (0.13)	0.70 (0.09)	0.87 (0.08)	0.35 (0.12)*	0.78 (0.08)
#.3	0.94 (0.04)	0.99 (0.04)	0.73 (0.11)	0.72 (0.11)	0.82 (0.09)	0.84 (0.09)	0.39 (0.12)*	0.73 (0.11)
#.4	0.74 (0.07)			0.88 (0.07)	0.68 (0.09)	0.83 (0.08)	0.61 (0.09)	0.64 (0.11)
#.5	0.83 (0.09)			0.57 (0.13)	0.82 (0.09)	0.75 (0.1)	0.88 (0.08)	0.87 (0.07)
#.6	0.73 (0.09)			0.88 (0.1)	0.70 (0.12)		0.71 (0.11)	0.64 (0.08)
#.7	0.70 (0.10)			0.69 (0.07)			0.39 (0.14)*	0.80 (0.07)
#.8				0.63 (0.09)			0.72 (0.11)	
#.9							0.73 (0.12)	

Source: European Commission Joint Research Centre; WJP Rule of Law Index 2011

Notes: (1) Numbers represent the kernel estimates of the Pearson correlation ratio ( $\eta^2$ ). (2) Bootstrap confidence intervals are given in parenthesis. (3) Sub-factors that have much lower contribution to the variance of the relevant Factor scores than the equal weighting expectation are marked with an asterisk. (4) F.1: Limited Government Powers, F.2: Absence of Corruption, F.3: Order and Security, F.4: Fundamental Rights, F.5: Open Government, F.6: Effective Regulatory Enforcement, F.7: Access to Civil Justice, F.8: Effective Criminal Justice

obtained if a given sub-factor could be fixed'. As discussed in Paruolo *et al.*, 2011, we can take this as a measure of importance; 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 factor, though some sub-factors are slightly more important than others (see Table 1). However, for the *Access to Civil Justice*, one could question the contribution of sub-factors 7.2, 7.3 and 7.7 compared to the remaining sub-factors on the basis of their lower effective weight. The issue is somewhat more serious for *Order and Security* where sub-factor 3.2 (*civil conflict is effectively limited*) is half as important as the other two. The reason is that 52 out of 66 countries do have civil conflict effectively limited and hence they all receive a score of 1.0 in this sub-factor. Consequently, sub-factor 3.2 has no discriminating power over those countries. Yet, sub-factor 3.2 becomes important and placed on equal footing with the other two sub-factors when it comes to the remaining 14 countries where civil conflicts exist. In order for sub-factor 3.2 to become as important as the other two for the entire set of countries, the original weights should be changed from 1,1,1, to 1, 2.5, 1 (in that case all  $S_i$  values will be between 0.60 and 0.70).

## Impact of modeling assumptions on the WJP Rule of Law Index results

Every dimension in the WJP Rule of Law Index is 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 robustness analysis is to assess to what extent 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. 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 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; on the other hand, the need to respect the rationale of the WJP that the sub-factors are equally important when calculating a dimension. Given these considerations, limit values of uncertainty intervals have been defined as shown in Table 2.

The multi-modeling approach involved combinations of the remaining two key assumptions on the 'no imputation' of missing data and the aggregation formula within a factor. The WJP calculated sub-factor scores using only available information for each country<sup>4</sup>. This choice (often termed as 'no imputation') was confronted with the application of the hot-deck imputation method<sup>5</sup>. 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 inherent theoretical inconsistencies lined to their fully compensatory nature, in which a comparative advantage of a few variables can compensate a comparative disadvantage of many variables. Hence, we considered the geometric average instead, which is a partially compensatory approach.<sup>6</sup> Consequently, we tested four models based on the combination of no imputation versus hot-deck and arithmetic versus geometric average. Combined with the 1,000

**Table 2. Uncertainties simulated in the WJP Rule of Law Index**

I. Uncertainty in the imputed values			
	Reference method	Alternative method	
	no imputation	hot-deck	
II. Uncertainty in the weights			
WJP Rule of Law Index 2011	Reference value	Uncertainty range (± 25% of reference value)	
Factor 1: Limited Government Powers (6)	0.167	0.125	0.208
Factor 2: Absence of Corruption (3)	0.333	0.250	0.417
Factor 3: Order and Security (3)	0.333	0.250	0.417
Factor 4: Fundamental Rights (8)	0.125	0.094	0.156
Factor 5: Open Government (6)	0.167	0.125	0.208
Factor 6: Effective Regulatory Enforcement (5)	0.200	0.150	0.250
Factor 7: Access to Civil Justice (8)	0.125	0.094	0.156
Factor 8: Effective Criminal Justice (7)	0.143	0.107	0.179
III. Uncertainty in the aggregation formula			
	Reference method	Alternative method	
	arithmetic average	geometric average	

*Source: European Commission Joint Research Centre; WJP Rule of Law Index 2011*

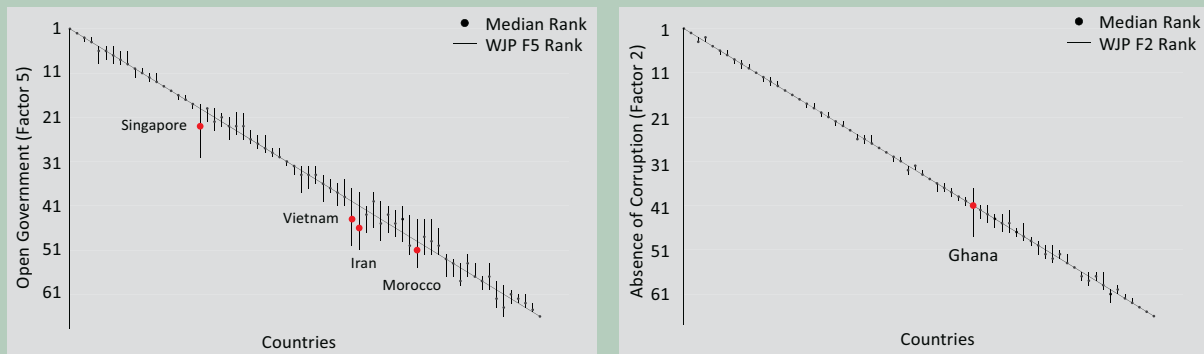
Notes: Number of sub-factors underlying each factor are given in parenthesis.

simulations per model to account for the uncertainty in the weights across the sub-factors, we carried out altogether 4,000 simulations.

The main results of the robustness analysis are provided in Figure 1, which shows median ranks and 90% intervals computed across the 4,000 Monte Carlo simulations for *Absence of Corruption* (F.2, one of the most robust dimensions) and for *Open Government* (F.5, one of the least robust dimensions). Countries are ordered from best to worst 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. Ranks in all eight factors are very robust to the modeling assumptions: 90 percent of the countries shift with respect to the simulated median less than ± 1 position in *Limited Government Powers* (F.1), *Absence of Corruption* (F.2), *Fundamental Rights* (F.4) and *Effective Regulatory Enforcement* (F.6); less than ± 2 positions in *Access to Civil Justice* (F.7) and *Effective Criminal Justice* (F.8); less than ± 3 positions in *Order and Security* (F.3) and *Open Government* (F.5). The fact

**Figure 1: Robustness analysis (WJP factor ranks vs. median rank, 90% intervals)**



*Source: European Commission Joint Research Centre; WJP Rule of Law Index 2011*

Notes: Countries with wide intervals –more than 10 positions– across 4,000 simulations related to estimation of missing data, weighting and aggregation formula are flagged.

that *Absence of Corruption* (F.2) 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) is all the more noteworthy given its potential inclusion in the Corruption Perception Index of Transparency International, to describe perception of corruption in the public sector and among politicians. Simulated 90% intervals across 4,000 Monte Carlo runs are narrow enough for most countries (less than 4 positions in 75% of the cases) to allow for meaningful inferences to be drawn. Exceptionally, few countries have relatively wide intervals (roughly 10–16 positions): China and Liberia on F.1, Ghana on F.2, Bangladesh on F.3, Singapore on F.4, Iran, Morocco, Singapore and Vietnam on F.5, Jamaica on F.6, and no country on F.7 or F.8. 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). Although these few cases are not a worrisome concern in the context of rule of law, they have been flagged herein as part of the sensitivity 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 assumptions made during the development phase.

Overall, the JRC analysis suggests that the conceptualized multi-level structure of the WJP Rule of Law Index is statistically coherent and none of the eight dimensions is dominated by any of its underlying sub-factors. Country ranks across the eight dimensions are also fairly robust to methodological changes related to the estimation of missing data, weighting or aggregation rule (less than  $\pm 1$  position shift in 90% of the cases).

## 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.

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

Paruolo, P., Saltelli, A., Saisana, M., Ratings and rankings: Voodoo or Science? *Journal of the Royal Statistical Society A* (submitted, 2011).

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.

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

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.

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

<sup>2</sup> 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 (66 countries).

<sup>3</sup> In the WJP Rule of Law Index, 'factors' are equivalent to dimensions and 'sub-factors' to sub-dimensions.

<sup>4</sup> Note that here 'no imputation' is equivalent to replacing missing values with the average of the available data within each sub-factor.

<sup>5</sup> The 'hot-deck method' (also termed 'nearest neighbour method') involves substituting missing values for a given country with available data from 'similar' countries, similarity being measured by a certain distance (Little and Rubin, 2002). For the WJP factors, after cross-validation, we selected Manhattan distance and three nearest neighbours.

<sup>6</sup> 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.