The World Justice Project

The World Justice Project | Rule of Law Index®

2012

Mark David Agrast
Juan Carlos Botero
Joel Martinez
Alejandro Ponce
Christine S. Pratt

The World Justice Project | Rule of Law Index®

2012

Mark David Agrast
Juan Carlos Botero
Joel Martinez
Alejandro Ponce
Christine S. Pratt

With the collaboration of:
Kelly Roberts

The *WJP Rule of Law Index® 2012* report was made possible by generous support from:

William H. Neukom & Neukom Family Foundation and the Bill & Melinda Gates Foundation.

And from the Allen & Overy Foundation; Allen & Overy LLP; American Bar Association; American Bar Association Criminal Justice Section; American Bar Association Health Law Section; American Bar Association Judicial Division; American Bar Association Section of Administrative Law and Regulatory Practice; American Bar Association Section of Antitrust Law; American Bar Association Business Law Section; American Bar Association Section of Dispute Resolution; American Bar Association Section of Environment, Energy, and Resources; American Bar Association Section of Individual Rights and Responsibilities; American Bar Association Section of Intellectual Property Law; American Bar Association Section of International Law; American Bar Association Section of Labor and Employment Law; American Bar Association Section of Litigation; American Bar Association Section of Real Property, Trust and Estate Law; American Bar Association Section of State and Local Government Law; American Bar Association Section of Taxation; Anonymous; Boies, Schiller & Flexner, LLP; The Boeing Company; Carnegie Corporation of New York; Chase Family Philanthropic Fund; Cochingyan & Peralta Law Offices; Drinker Biddle & Reath LLP; E. I. du Pont de Nemours and Company; Eastminster Presbyterian Church; The Edward John and Patricia Rosenwald Foundation; The Ewing Marion Kauffman Foundation; Ford Foundation; Fulbright & Jaworski LLP; Garrigues LLP; GE Foundation; General Electric Company; Gómez-Acebo & Pombo; Google, Inc.; Gordon and Betty Moore Foundation; Haynes and Boone, LLP; Hewlett-Packard Company; Holland & Knight LLP; Hunton & Williams; Intel Corporation; Irish Aid; Johnson & Johnson; Judson Family Fund at The Seattle Foundation; K&L Gates; LexisNexis; Major, Lindsey & Africa; Mason, Hayes+Curran; McKinsey & Company, Inc.; Merck & Co., Inc.; Microsoft Corporation; National Endowment for Democracy; Nelson Mullins Riley & Scarborough LLP; Oak Foundation; PepsiCo; Roca Junyent; Society of the Cincinnati; Sullivan & Cromwell LLP; SyCip Salazar Hernandez & Gatmaitan; Texas Instruments, Inc.; Troutman Sanders LLP; Turner Freeman Lawyers; U.S. Chamber of Commerce & Related Entities; Uría Menéndez; Viacom International, Inc.; Wal-Mart Stores, Inc.; Welsh, Carson, Anderson & Stowe; White & Case LLP; William and Flora Hewlett Foundation; Winston & Strawn LLP; and individual supporters listed in the last section of this report.

ISBN (print version): 978-0-9882846-2-3 **ISBN (online version): 978-0-9882846-3-0**

Graphic design: Joshua Steele.

Suggested citation: Agrast, M., Botero, J., Martinez, J., Ponce, A., & Pratt, C. *WJP Rule of Law Index®* 2012. Washington, D.C.: The World Justice Project.

Contents

Part III: JRC audit on the 2012 WJP Rule of Law Index

Statistical Audit

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 2012 WJP Rule of Law Index is statistically coherent and balanced (i.e., none of the eight dimensions is dominated by an underlying component). Furthermore, the analysis has offered statistical justification for the equal weights and the use of arithmetic averaging at the various levels of aggregation. 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 ± 3 positions 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 choices on a country's performance are useful steps: they add to the transparency and reliability of the Index and to build confidence in the narratives supported by the measure. Modelling 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 single set of numbers.

The Econometrics and Applied Statistics Unit at the European Commission Joint Research Centre in Ispra (Italy) has undertaken for a third consecutive year, upon invitation of the WJP, a thorough statistical assessment of the *Index*[1]. Fine-tuning suggestions made by the JRC for the previous two releases of the *Index* were already taken on board by the WJP. However, due to some re-structuring of the framework from 46 to 44[2] sub-factors and from 479 to 516 survey questions, the WJP requested an audit of the Index for a third time. The WJP Rule of Law Index was assessed along two main avenues: the conceptual and statistical coherence of the structure, and the impact of key modeling choices on its 2012 WJR Rule of Law scores and ranks.

CONCEPTUAL AND STATISTICAL COHERENCE IN THE WJP RULE OF LAW FRAMEWORK

Country data delivered to the JRC were average scores across academics or individuals along 516 survey questions (henceforth variables) for 97 countries. These variables are not affected by outliers or skewed

1 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/

2 A total of only 44 sub-factors accounts for the fact that neither sub-factor 1.1, 9.1, 9.2 nor 9.3 are covered in the 2012 Index.

distributions[3], except for 13 variables spread across five 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. The 2012 dataset is characterized by excellent data coverage (96% in a matrix of 516 variables × 97 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 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.

Principal component analysis (PCA) was used to assess to which extent the conceptual framework is confirmed by statistical approaches and to identify eventual pitfalls. The analysis confirms the 2012 WJP Rule of Law Index structure, as within each of the eight dimensions the first latent factor captures between 58% up to 87% of the variance (best result for the dimension on *Absence of Corruption*). 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, the conceptual grouping of sub-factors into dimensions is statistically supported by the data. Finally, the eight dimensions share a single latent factor that captures 81% 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 to 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 the eight dimension scores that would be obtained if a given sub-factor could be fixed'. As discussed in Paruolo *et al.*, 2012, we can take this as a measure of importance[5]; thus if sub-factors are

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 (97 countries).

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

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.

Table 4: Importance measures (variance-based) for the 44 sub-factors in the eight dimensions of the 2012 WJP Rule of Law Index

Sub-factor	D1	D2	D3	D4	D5	D6	D7	D8
#.1		0.95 [0.93, 0.96]	0.64 [0.63, 0.72]	0.51* [0.51, 0.56]	0.76 [0.73, 0.80]	0.81 [0.80, 0.84]	0.58 [0.56, 0.60]	0.69 [0.62, 0.74]
#.2	0.87 [0.83, 0.90]	0.90 [0.85, 0.91]	0.42* [0.42, 0.44]	0.84 [0.82, 0.89]	0.81 [0.80, 0.87]	0.91 [0.88, 0.91]	0.55 [0.55, 0.66]	0.76 [0.76, 0.82]
#.3	0.92 [0.89, 0.92]	0.91 [0.88, 0.93]	0.62 [0.62, 0.71]	0.72 [0.72, 0.78]	0.73 [0.72, 0.83]	0.74 [0.71, 0.8]	0.82 [0.79, 0.84]	0.80 [0.78, 0.86]
#.4	0.81 [0.80, 0.84]	0.84 [0.81, 0.87]		0.79 [0.75, 0.84]	0.81 [0.79, 0.86]	0.82 [0.80, 0.85]	0.64 [0.58, 0.71]	0.69 [0.69, 0.78]
#.5	0.72 [0.71, 0.77]					0.74 [0.71, 0.82]	0.43* [0.43, 0.52]	0.87 [0.86, 0.90]
#.6	0.80 [0.75, 0.85]						0.80 [0.77, 0.82]	0.60 [0.47, 0.69]
#.7	0.83 [0.80, 0.86]						0.60 [0.59, 0.68]	0.86 [0.85, 0.87]
#.8				0.65 [0.65, 0.69]				

Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law 2012
Notes: (1) Numbers represent the kernel estimates of the Pearson correlation ratio (n^2), as in Paruolo et al., 2012. Min-max estimates for the n^2 derive from the choice of the smoothing parameter. (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: Limited Government Powers, D2: Absence of Corruption, D3 Order and Security, D4: Fundamental Rights, D5: Open Government, D6: Regulatory Enforcement, D7: Civil Justice, D8: Criminal Justice.

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 (see Table 4 "Table 4: Importance measures (variance-based) for the 44 sub-factors in the eight dimensions of the 2012 WJP Rule of Law Index" on page 11). Although still acceptable, the worst results for this kind of coherence analysis are: under *Fundamental Rights* dimension (D4), the contribution of the sub-factor 4.1 (*equal treatment and absence of discrimination*) compared to the remaining sub-factors on the basis of the lower effective weight. Similarly, sub-factors 3.2 (*civil conflict is effectively limited*) and sub-factor 7.5 (*civil justice is not subject to unreasonable delays*) have a lower contribution to the variance of the respective dimension compared to the other sub-factors in those dimensions. All together the degree of coherence of this version of the index is remarkable, i.e. all dimensions look balanced and coherent.

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 uncertainty 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

Table 5: Uncertainty parameters (missing values, weights and aggregation function)

	Reference	Alternative
I. Uncertainty in the treatment of missing data	No estimation of missing data	Expectation Maximization (EM)
II. Uncertainty in the aggregation function	Arithmetic average	Geometric average
III. Uncertainty intervals for the sub-factor weights	Reference value for the weight	Distribution assigned for uncertainty analysis
1: Limited Government Powers (# 6 sub-factors)	0.167	U[0.125, 0.208]
2: Absence of Corruption (#4 sub-factors)	0.250	U[0.187, 0.312]
3: Order and Security (#3 sub-factors)	0.333	U[0.250, 0.417]
4: Fundamental Rights (#8 sub-factors)	0.125	U[0.094, 0.156]
5: Open Government (#4 sub-factors)	0.250	U[0.187, 0.312]
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 2012

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-modelling 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 and equally important when calculating a dimension. Given these considerations, limit values of uncertainty intervals have been defined as shown in Table 5.

The multi-modelling approach involved combinations of the remaining two key assumptions on the 'no imputation' of missing data and the aggregation formula within a dimension. The WJP calculated sub-factor scores using only available information for each country[6]. This choice (often termed as 'no imputation') was confronted with the application of the expectation-maximization method for the estimation of the missing data[7]. 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

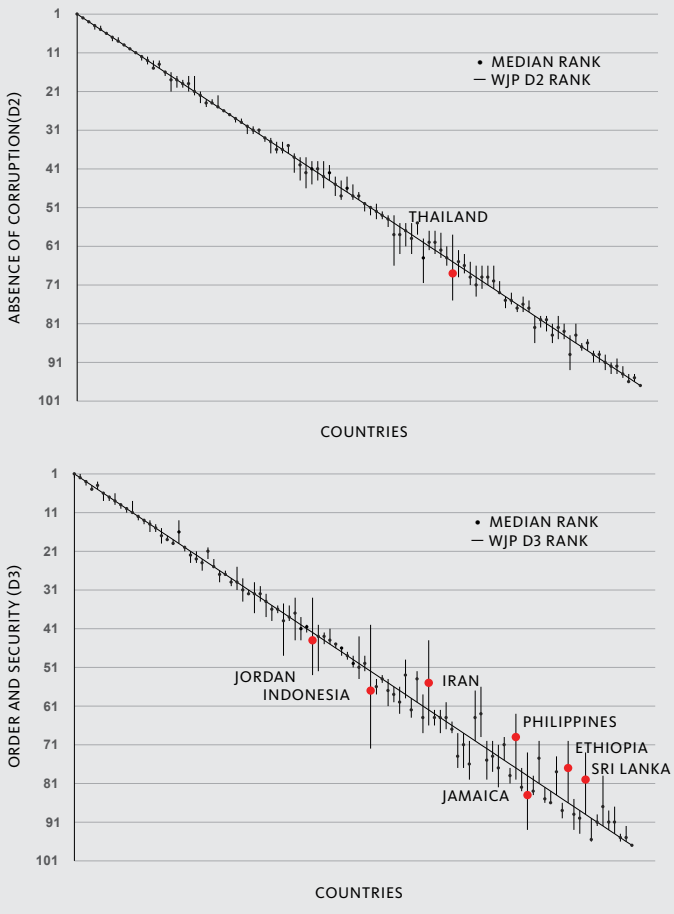
6 Note that here 'no imputation' is equivalent to replacing missing values with the average of the available data within each sub-factor.

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.

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. This offsetting might not be always desirable when dealing with fundamental aspects of a concept such as rule of law. Hence, we considered the geometric average instead, which is a partially compensatory approach[8]. Consequently, we tested 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.

The main results of the uncertainty 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* (D2, one of the most robust dimensions) and for *Order and Security* (D3, 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 dimensions are very robust to the modeling assumptions: 90 percent of the countries shift with respect to the simulated median less than ± 2 positions in *Limited Government Powers* (D1) and *Absence of Corruption* (D2); less than ± 3 positions in *Regulatory Enforcement* (D6), *Civil Justice* (D7) and *Criminal Justice* (D8); less than ± 4 positions in *Fundamental Rights* (D4); less than ± 5 positions in *Open Government* (D5); less than ± 6 positions in *Order and Security* (D3).

Figure 9: Uncertainty analysis
(WJP dimension ranks vs. median rank, 90% intervals)



Source: Saisana and Saltelli, European Commission Joint Research Centre; WJP Rule of Law 2012
Notes: Countries with wide intervals –more than 15 positions– across 4,000 simulations related to estimation of missing data, weighting and aggregation formula are flagged.

The fact that the dimension on *Absence of Corruption* (D2) is one of the most robust in the WJP Rule of Law Index with respect to modeling assumptions and also very coherent (as discussed in the previous section, see Table 4)

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 close to zero values biasing the geometric average, we re-scaled linearly the sub-factors scores to a minimum of 0.1.

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).

Overall across all 97 countries and eight dimensions of the rule of law, there is an absolute shift of less than 3 positions with respect to the simulated median rank in 90% of the cases. Note that in the 2011 release of the index (66 countries) the respective shift was merely 1 position for 90% of the cases. This should not be interpreted as the 2012 being less robust given the higher number of countries included this year (97 in 2012, over 66 in 2011).

Simulated 90% intervals across 4,000 Monte Carlo runs are narrow enough for most countries (less than 6 positions in 75% of the cases) to allow for meaningful inferences to be drawn. Few countries have relatively wide intervals (more than 15 positions): none on D1; Thailand on D2; Ethiopia, Indonesia, Iran, Jamaica, Jordan, Philippines and Sri Lanka on D3; none on D4; Albania, China, Iran, Madagascar, Malaysia, Senegal, Sri Lanka, and UAE on D5; none on D6; Croatia and Madagascar on D7; Peru on 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.

CONCLUSION

The JRC analysis suggests that the conceptualized multi-level structure of the 2012 WJP Rule of Law Index is statistically coherent and balanced (i.e., none of the eight dimensions is dominated by an underlying component). Furthermore, the analysis has offered statistical justification for the equal weights and the use of arithmetic averaging at the various levels of aggregation – which should not be taken for granted when linear aggregation is concerned. 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 ± 3 positions shift in 90% of the cases). A hypothetical aggregated Rule of Law Index would also appear statistically justified given the data. Finally, the fact that the dimension on Absence of Corruption is especially coherent and robust in the WJP Rule of Law Index is noteworthy given its inclusion in the Corruption Perception Index of Transparency International.

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. 2012. Ratings and rankings: Voodoo or Science? *Journal of the Royal Statistical Society A,* 176 (2), 1-26.

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. Rickety 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.