2018
IBRAHIM INDEX OF AFRICAN GOVERNANCE

METHODOLOGY

MO IBRAHIM FOUNDATION

CONTENTS

INTRODUCTION TO THE IIAG

The Ibrahim Index of African Governance (IIAG) provides an annual statistical assessment of governance performance in 54 African countries.

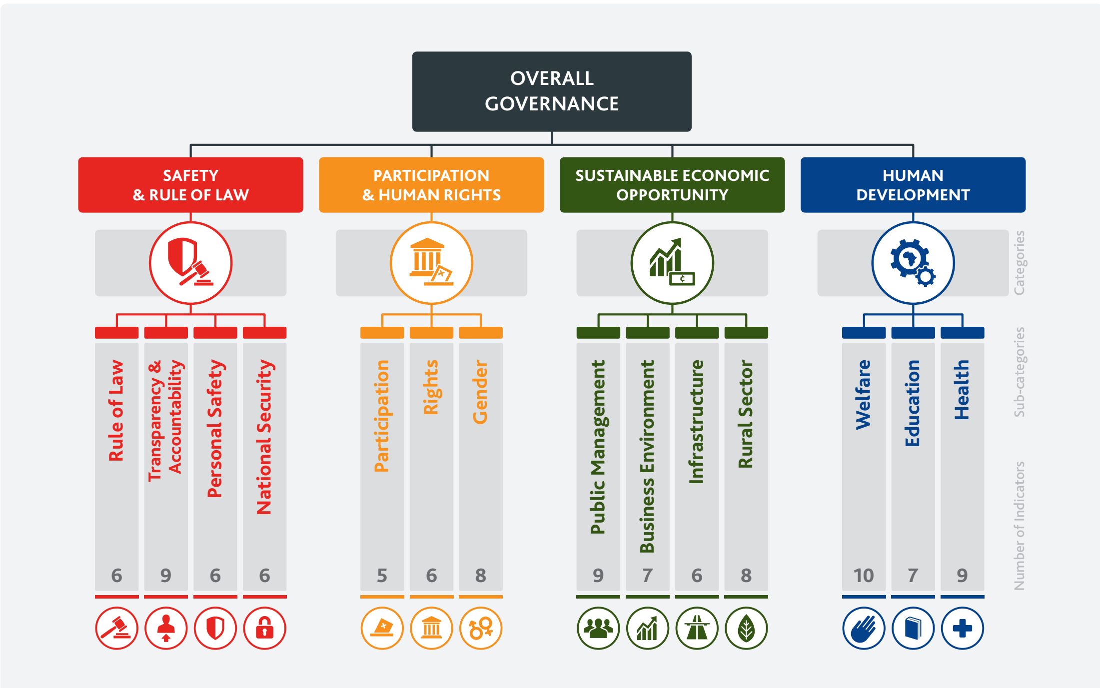
Data from diverse sources are combined into composite governance scores – an *Overall Governance* score, categories, sub-categories and some composite indicators, which all measure different aspects of governance. Every component of the IIAG is comparable for the time period 2008-2017 enabling analysis of national and regional trends over time. The entire Index time series is updated on an annual basis to ensure that each new IIAG provides the most accurate data available.

This document contains a detailed explanation of how the IIAG is calculated. The methodology is simple and transparent. All of the underlying data and information used in the construction of the IIAG are freely available and transparently published by the Mo Ibrahim Foundation, alongside the scores and uncertainty estimates. Given the inherently unobservable nature of governance, all measures are made with error and users of the Index are encouraged to use the confidence intervals whilst interpreting the scores. This document will outline how the Foundation conceptualises governance, assumptions made, how measures are selected for inclusion, how the IIAG is calculated and how we estimate uncertainty around the scores.

MEASURING GOVERNANCE

The Mo Ibrahim Foundation (MIF) defines governance as the provision of the political, social and economic goods and services that every citizen has the right to expect from their state, and that a state has the responsibility to deliver to its citizens. The IIAG is concerned with operationalising these public goods through outputs and outcomes of policy, as opposed to inputs or *de jure* indicators. The framework of the IIAG has been constructed to reflect this definition and has four main pillars of governance: *Safety & Rule of Law*, *Participation & Human Rights*, *Sustainable Economic Opportunity* and *Human Development*. These categories are comprised of 14 sub-categories, as shown in Table 1.

Table 1. 2018 IIAG Structure



Each construct is operationalised as a composite indicator of its sub-components. The *Overall Governance* score is constructed by calculating an unweighted average of its underlying four categories; these categories are constructed by calculating an unweighted average of their respective sub-categories. Likewise, sub-category scores are the result of aggregating the scores of all their underlying indicators, which are formed by 191 variables. A variable is anything that can be constituted as a raw data from source. There are 191 variables collected to calculate the 2018 IIAG, which sit under 102 indicators, 48 of which are clustered and 54 standalone indicators. In total, the 2018 Index contains 273 measures of governance (taking into account the variables collected from source as well as every composite score provided in the IIAG dataset). The data comes from 35 data providers, a mix of qualitative and quantitative assessments.

TECHNICAL ASSUMPTIONS

The central assumption is that each indicator score is the true value of its respective sub-category score plus some error. Each constituent indicator measures something specific – for example *Primary School Completion* – but when included in a sub-category it is expected to measure a broader concept; in this case whether a country has a comprehensive education system. In countries which have good primary school education, secondary and tertiary education may be less developed, and therefore primary school completion rates may overestimate the adequacy of education provided in a country. It is this misestimation which is captured in the error term, as discussed further in the section on uncertainty estimates.

The IIAG makes the assumption of orthogonality across indicator errors. This is illustrated in the *Online Public Services* and *Diversification of Exports* indicators from the *Public Management* sub-category. Both indicators misestimate *Public Management* by some amount, but because the indicators come from two very different sources – the E-Government Development Index from the United Nations Department of Economic and Social Affairs and the African Economic Outlook produced by the African Development Bank, Organisation for Economic Co-operation and Development and United Nations Development Programme - there is no reason to suppose that when one indicator overestimates *Public Management* the other indicator will do the same. It is assumed that the errors are not related and will cancel each other out when the indicators are aggregated. The more indicators are added the more likely it is that the errors sum to zero. Any correlation between indicators arises as a result of a latent governance dimension. This misestimation is elaborated in greater detail in the section on uncertainty estimates.

MEASUREMENT SELECTION & CRITERIA FOR INCLUSION

To be included in the IIAG, a variable has to be a suitable proxy for governance and should measure outputs and outcomes of governance, not inputs or de jure measures. For example, to measure the level of education in a country, we use the indicator *Education Quality*, as opposed to expenditures made by a government on education. Further considerations around inclusion of a variable include their analytical soundness, timeliness and accessibility.

For each dataset under consideration, missing data is a large issue. In particular, the number of countries covered by each dataset in a certain year; the number of years covered; the periodicity of the data and the most recent year of published data are all parameters which impact the precision of the composite scores.

Given these considerations, to be included in the IIAG, a variable must have at least two years' worth of data since the beginning of the time series (2008) for at least 33 countries and the latest data point for these 33 countries must exist within the last three years. Furthermore, in order to differentiate between scores, numerical granularity is taken into consideration, with all the measures being on a four-point scale or more.

When a variable is deemed to be suitable for inclusion, it is assigned to the sub-category in which it sits best conceptually. As dimensions of governance are not independent and variables may be deemed to be suitable for multiple sub-categories, this process requires consultation with the IIAG Advisory Council, a group of expert advisors.

OUTLIERS & TREATMENT

In some instances, a variable includes observations which lie far away from the mass of the rest of the distribution. Including these extreme observations in the IIAG would bias the measure scores as after normalisation the outliers would make the range skewed and differentiation between most of the countries' scores would be difficult. To prevent this, the raw data is analysed to determine whether any of the variables require treatment to address outliers.

Outlier diagnostics identify outliers in the raw data, for each variable independently. All of these diagnostics are based on Tukey's method, which measures the distance of extreme observations from an upper and lower threshold. These thresholds are calculated using the interquartile range (IQR), the 1st and 3rd quartile and a factor. If a point lies above the upper threshold or below the lower threshold is considered an outlier.

$$\text{Interquartile range (IQR)} = 3^{rd}\ \text{quartile} - 1^{st}\ \text{quartile}$$
$$\text{Upper threshold} = 3^{rd}\ \text{quartile} + IQR * \text{factor}$$
$$\text{Lower threshold(IQR)} = 1^{st}\ \text{quartile} - IQR * \text{factor}$$

In some cases, the decision to treat outliers was clear, such as *Absence of Internally Displaced Persons*, in which the most extreme outlier was more than twenty times the distance from the centre of the distribution. In other instances, which may have been just on the threshold of the trimming diagnostics, further inspection of the distribution, knowledge around the phenomenon measured and considerations around information lost if trimmed, factored into the decision around whether to treat a variable or not.

In the 2018 IIAG, seven variables were treated for outliers: *Absence of Riots & Protests*, *Absence of Government Violence against Civilians*, *Absence of Violence against Civilians by Non-state Actors*, *Absence of Internally Displaced Persons*, *Absence of Refugees* and *Absence of Malaria*.

These variables were filtered according to the following scheme: the trimmed mean and trimmed standard deviation of the measure were computed on the central 95% of the distribution; i.e. the bottom and top 2.5% were not used to compute the mean and standard deviation. All observations further than 3 trimmed standard deviations away from the trimmed mean were replaced by the trimmed mean plus 3.1 times the trimmed standard deviation, if they were in the right-hand tail, and replaced by the trimmed mean minus 3.1 times the trimmed standard deviation, if they were in the left-hand tail.

DEALING WITH MISSING DATA

The majority of the variables included in the IIAG have a degree of missing data over the time series (2008-2017). To ensure continuity between composite scores over time it is necessary to impute values for these years.

In order to determine the most suitable method of imputation for the IIAG, simulation experiments were conducted, in which a proportion of the data were deleted using various missingness mechanisms. The deleted data were then imputed using a range of methods.

From the point of view of accuracy, precision and the amount of missingness remaining after imputation, it was determined that the best method of imputation was linear interpolation for missing data which are in the interior of the time series. The interior missing values are replaced with numbers incrementally higher or lower than the neighbouring data points.

For missing data points that are located in the exterior of the available time series, data for a country in a previous or following year was deemed to be the best proxy available to measure governance in the given year. Hence, the exterior missing values are replaced using the closest data point from source (last value carried forward - LVCF - or first value carried backward - FVCB).

As an example, Country A has data missing for Year 1, Year 4 and Year 5. The first row of Table 2 contains raw data. For the interior missing values in Year 4 and Year 5, linear interpolation is used to obtain the value, as shown in the second row of Table 2. For the exterior missing values in Year 1, the missing data is imputed using FVCB to obtain the data as shown in the third row.

Table 2. Imputation example for Country A

Country A	2008	2009	2010	2011	2012	2013
Raw data from source		85.1	84.8			82.8
Between data points		85.1	84.8	*84.1*	*83.5*	82.8
Outside data points	*85.1*	85.1	84.8	*84.1*	*83.5*	82.8

Let x_t be the raw data value of a variable for a country in year t. If there is no data for year t_2, and there are data for the years t_1 and t_3, whereby t_1 and t_3 are the closest such years to t_2 with the property that $t_1<t_2<t_3$. Linear interpolation can be illustrated as follows.

$$x_2 = x_1 + (t_2 - t_1) \times \frac{x_3 - x_1}{t_3 - t_1}$$

Using the above formula in the previous example, a missing value in 2011 would be imputed as follows using linear interpolation:

$$84.8 + (2011 - 2010) * \frac{(82.8 - 84.8)}{2013 - 2010} = 84.1$$

Exceptions

There are exceptions for when imputation is not applied. For countries which have missing data for the entire time series for a variable, no imputation is carried out. Similarly, for countries which only have one data point across the time series, imputing this point across the whole time series using LVCF/FVCB would be inappropriate due to the inaccuracies incurred.

NORMALISATION

Data used in the construction of the IIAG are diverse and include both quantitative and qualitative assessments provided by a range of data providing institutions. At source, the variables collected are produced on different scales and can also have different polarities (e.g. in raw data terms, while for *Access to Justice* higher score is better, for *Absence of Government Involvement in Armed Conflict* higher score is worse). In order for them to being meaningfully combined and compared, raw data are standardised before being included in the IIAG.

The data points for each variable are transformed using the min-max normalisation method. This method performs an order preserving linear transformation of the data, in which the maximum value in the raw data becomes the highest score and the minimum value in the raw data becomes the lowest score for each variable.

The min-max normalisation method subtracts from the actual value the minimum value of the entire raw dataset (2008-2017) in the specific variable and divides by the range of the variable values. The scores for each country's value in a particular year where the polarity of the raw data is 'higher score is better' are transformed as follows:

$$x = 100 \times \frac{\text{actual value} - \text{minimum raw value}}{\text{maximum raw value} - \text{minimum raw value}}$$

If the polarity of the raw data is 'higher score is worse', the scores must be inverted. Therefore, the previous formula must be altered as follows:

$$x = 100 - 100 \times \frac{\text{actual value} - \text{minimum raw value}}{\text{maximum raw value} - \text{minimum raw value}}$$

This method of standardisation allows all scores to be published in common units and within the same bounds of 0.0-100.0, in which 100.0 reflects the best performance in Africa across the ten years. 100.0 does not imply perfect governance but indicates the best score in Africa according to the raw data. Thus, meaningful comparisons between variables and countries can be made. Other advantages of this method include not being affected by skewed values and not making any assumptions about the distribution of the data.

AGGREGATION & WEIGHTING

The IIAG uses linear, additive aggregation and weights each sub-component equally within its dimension. The IIAG uses linear, additive aggregation and weights each sub-component equally within its dimension. While there are a number of different types of aggregation methods with respective pros and cons, there is no set standard for aggregation in composite indices. The linear aggregation method has advantages in its simplicity, transparency and accessibility. The decision to weight the four overarching categories equally in the IIAG was taken based on the judgment that the four governance dimensions of the IIAG – *Safety & Rule of Law*, *Participation & Human Rights*, *Sustainable Economic Opportunity* and *Human Development*- are of equal importance in measuring governance.

While the weight of the categories is equal in the *Overall Governance* composite score, sub-categories have different implicit weighting as a result of the structure of the IIAG. Each category is comprised of a different number of sub-categories, which in turn include a diverse number of indicators. For example, *Human Development* is comprised of three underlying sub-categories (*Welfare*, *Education* and *Health*), while *Sustainable Economic Opportunity* is an aggregate of four sub-categories (*Public Management*, *Business Environment*, *Infrastructure* and *Rural Sector*).

This effectively means that while Welfare has an overall weighting of 8.3 percent in the *Overall Governance* score, *Public Management* carries a smaller weighting of 6.3 percent.

The number of sub-categories within a category is based on the theoretical framework and data that are available to proxy the dimension. *Human Development* is conceptually defined as the provision of a decent standard of living, comprehensive education and a healthy life, reflected in its three underlying sub-categories. *Sustainable Economic Opportunity* captures the provision of the conditions necessary to pursue economic opportunities that contribute to a prosperous and equitable society, reflected in its four sub-categories.

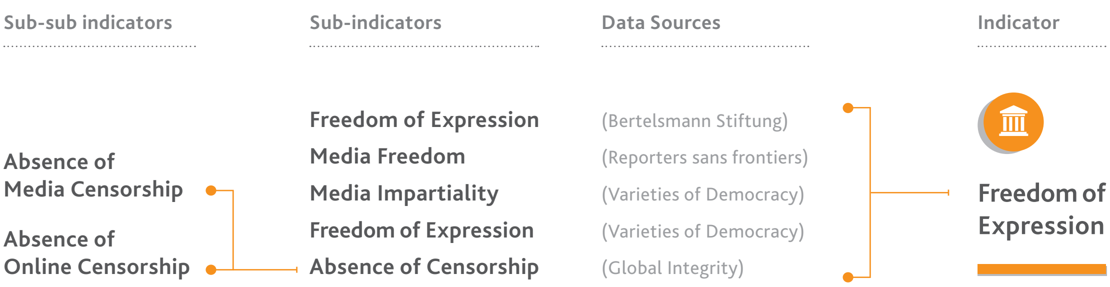
CLUSTERING

The underlying 191 variables included in the IIAG are constituents of 102 indicators, 48 of which are clustered indicators. A clustered indicator or a sub-indicator is composed of a number of underlying variables which capture the same dimension.

The data included in the IIAG comes from 35 different institutions and are a mix of qualitative and quantitative assessments. The diversity of sources and types of data means the measurement error in the composite values is minimised.

A measure can be clustered as a result of data on the same issue being available from multiple sources or when a measure is deemed to be too narrow to include as a standalone indicator, it can be combined in a clustered indicator to measure a slightly broader concept.

For example, variables measuring narrow concepts such as *Absence of Media Censorship* and *Absence of Online Censorship*, from Global Integrity, are clustered together to create the *Absence of Censorship* sub-indicator. In turn this sub-indicator together with the sub-indicators *Media Freedom*, from Reportes San Frontières, *Media Impartiality*, from Varieties of Democracy, and *Freedom of Expression*, from both Bertelsmann Stiftung and Varieties of Democracy, are clustered together to create the indicator *Freedom of Expression*.



QUANTIFYING UNCERTAINTY

The Foundation publishes standard errors and confidence intervals alongside the *Overall Governance* and category scores to reflect degrees of uncertainty. The standard errors and confidence intervals allow users of the IIAG to discriminate, to a certain degree, between changes in the values of the IIAG that can be confidently treated as actual changes in the state of governance and changes that might be due to "noise", or are at least insufficiently sizeable to be able to ascribe a high likelihood to such change being statistically significant.

Users of the IIAG are encouraged to use these to interpret changes in country scores over time and differences between country scores. For example, score or rank comparisons between countries with overlapping confidence intervals should generally be avoided as they represent a statistical tie. Similarly, users are encouraged not to over interpret marginal differences in a country's score change and refer to the confidence intervals for statistically significant movements.

The uncertainty in the IIAG arises from measurement and missingness error. Two assumptions are made around the error. Firstly, each measure in the IIAG is a measurement of its true sub-category score plus some measurement error. Secondly, the imputation of missing values is made with some missingness error. These two errors are combined to obtain the uncertainty estimates for *Overall Governance* and category scores.

Measurement error

The measurement error calculates how well each measure reflects the true value of its sub-category score. All the measures in each category are bootstrapped to obtain new sub-category scores. Bootstrapping is a process of randomly re-sampling the original dataset to generate new datasets. The bootstrapped scores are then combined to provide new category estimates. This process is carried out a very large number of times to obtain a distribution of scores for all available, non-imputed data. The measurement error is calculated as the standard deviation of the distribution of all the bootstrapped values.

Missingness error

The missingness error calculates the uncertainty of the IIAG imputation method. As such, values which are missing at source are imputed based on the imputation methods described in the section on imputation. All of the missing values are deleted and re-imputed with a suitable replacement several times. Values are replaced based on their composite score and replaced by a data point generated from a normal distribution with mean and variance equal to its respective measure. The missingness error is calculated as the standard deviation of the distribution of all pseudo-imputations.

Standard errors

The standard errors are obtained by combining the measurement and missingness errors using Rubin's formulation. The formula below illustrates this, where *M* is the number of simulations of imputation, σ_{miss} is the standard deviation of the missingness errors and σ_{meas} is the standard deviation of the measurement errors.

$$\sqrt{\left((1+\frac{1}{M}) \times \sigma_{miss}^{2}\right) + \sigma_{meas}^{2}}$$

Confidence intervals

The standard errors are used to construct confidence intervals for the country scores. These confidence intervals allow users of the IIAG to discriminate between country scores and country trends over time. While the Foundation chooses to use 90% confidence intervals, other degrees of confidence (80%; 85%; 90% and 95%) are available on our website.

READING THE RESULTS

Historical revision

The IIAG is refined and revised on an annual basis to continually improve its measurement of governance. Improvements are a result of either methodological changes, or based on the inclusion of new data. Equally, if previously included measures undergo fundamental methodological changes or do not meet the criteria for inclusion it may be necessary to exclude them from future iterations. It is also necessary to update previously published data if retrospective revisions are made to data at source.

As a result of these changes, the IIAG is re-calculated every year. The retrospective revision means that no previous publications should be compared to the 2018 IIAG scores as differences may be a result of a change in framework or an update in data from source rather than due to a change in score. Score and rank comparisons between years should be made entirely within the 2018 IIAG.

Relativity

The IIAG country scores and ranks are all relative, taking into account a country's performance in relation to the other 54 African countries. This is a result of the normalisation procedure, which transforms the raw data into a scale of 0.0-100.0, whereby 100.0 is the best score. This means that a country's change in score and rank may be reflective of other countries performing better or worse. Users of the IIAG are encouraged to treat marginal differences in scores and ranks with caution and refer to the standard errors for statistically significant changes.

mo.ibrahim.foundation

/MoIbrahimFoundation

@Mo_IbrahimFdn #IIAG

moibrahimfoundation