2017

IBRAHIM INDEX OF AFRICAN GOVERNANCE

DETAILED METHODOLOGY

MO IBRAHIM FOUNDATION

CONTENTS

INTRODUCTION TO THE IIAG

Measuring the multiple dimensions of governance is necessary to provide a quantifiable tool to monitor governance performance and progress and support the development of effective and responsive policy solutions. The Ibrahim Index of African Governance (IIAG) provides an annual assessment of governance performance in every African country which is holistic, objective and transparent.

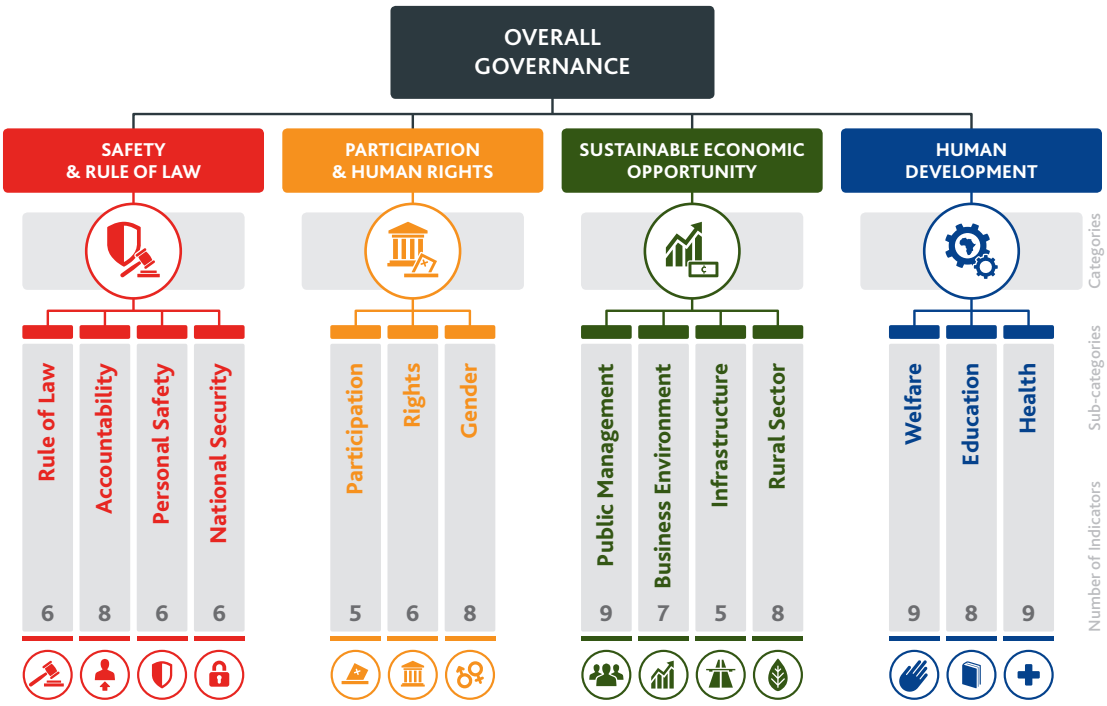
The entire Index time series is updated on an annual basis to ensure that each new IIAG provides the most accurate data available. Data from diverse sources are combined into composite governance scores – an *Overall Governance* score, categories, sub-categories and some composite indicators, which all measure different aspects of governance. Every component of the IIAG is comparable from the year 2000 enabling analysis of national and regional trends over time.

This document contains a detailed explanation of how the IIAG is calculated. The methodology is simple and transparent. All of the underlying data 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. 2017 IIAG Structure



As governance and its dimensions are not directly observable, 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 include 177 variables. A variable is anything that can be constituted as a raw data from source. There are 177 variables collected to calculate the 2017 IIAG, which sit under 100 indicators, 43 of which are clustered and 57 standalone indicators. In total, the 2017 Index contains 251 measures of governance (taking into account the variables collected from source as well as every composite score provided in the IIAG data set). The data comes from 36 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 *Governmental Statistical Capacity* and *Diversification* 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 Bulletin Board on Statistical Capacity from the World Bank 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. This assumption allows a distinction between noise and signal in the indicator values. 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. The MIF definition of governance is the provision of political, economic or social public goods which the state has the responsibility to deliver to its citizens and citizens have the right to expect from their state. The IIAG measures 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 *Literacy*, 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 data set under consideration, missing data is a large issue. In particular, the number of countries covered by each data set 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 (2000) 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

The variables included in the IIAG are published on different scales. 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 and mean that after normalisation the range would be 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. All of these diagnostics are based on Turkey's method, which measures the distance of extreme observations from the inter-quartile range. Thresholds for identifying outliers were developed based on these measures. In some instances, the case for treating outliers was clear, such as *Internally Displaced People*, 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 2017 IIAG, seven variables were treated for outliers: *Riots & Protests*, *Government Violence against Civilians*, Violence *by Non-state Actors*, *Internally Displaced People*, P*olitical Refugees*, *Budget Balance* and *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 (2000-2016). 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 (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	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
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

Linear interpolation can be illustrated as follows. Let x_{ti} be the raw data value of a variable for country i 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 $t_1<t_2<t_3$.

$$\hat{x}_{t_2 i} = x_{t_1 i} + (t_2 - t_1)\frac{x_{t_3 i} - x_{t_1 i}}{t_3 - t_1}$$

Using the above formula, an interior missing value in 2013 would be imputed as follows using linear interpolation:

$$83.1 + (2014 - 2013) \times \frac{(85.2 - 83.1)}{2014 - 2012} = 84.2$$

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 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 *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 data set (2000-2016) 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:

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

If the polarity of the raw data is 'higher score is worse', 'x'=100-100 x the stated formula. 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 is the best possible score. 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. 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 177 variables included in the IIAG are constituents of 100 indicators, 43 of which are clustered indicators. A clustered indicator is composed of a number of underlying variables which capture the same dimension. A sub-cluster is formed when a narrow issue is measured by the same or different sources. The data included in the IIAG comes from 36 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.

An indicator can be clustered as a result of data on the same issue being available from multiple sources. For example, political violence is measured by two different data sources included in the IIAG: Armed Conflict Location & Event Data Project (ACLED) and Political Terror Scale (PTS). To improve the accuracy of the indicator measurement and avoid overrepresenting the weight of each variable, these underlying variables are combined into a single clustered indicator.

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, the measures *IT Infrastructure* from the Economist Intelligence Unit, *Mobile Phone Subscribers*, *Household Computers* and *Household Internet Access*, all taken from the International Telecommunications Union, are included in the clustered indicator *Digital & IT Infrastructure*.

A sub-indicator captures narrow issues which are combined to measure a relatively broader issue. Concepts from the same source can be combined to measure a similar concept. For example, *Freedom of Association* is a sub-cluster of two measures from *Global Integrity*, *Freedom of Association* and *Trade Unions*, which are combined into a sub-indicator, before being clustered with *Freedom of Association & Assembly* from Bertelsmann Stiftung, to form the *Freedom of Association & Assembly* indicator.

A sub-indicator can also be the result of combining a narrow issue captured by two independent sources. *Women in Cabinet* from Global Integrity, and *Women in Ministerial Positions* from the World Bank, are combined to form the sub-cluster *Women in Ministerial Positions*, before being clustered with the sub-indicator *Women in Parliament* from the Inter-Parliamentary Union, to form the *Women's Political Participation* indicator. This indicator includes data from both quantitative and qualitative assessments to form a nuanced picture of women's participation in politics.

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.

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 data set to generate new data sets. 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.

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

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 2017 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 2017 IIAG

Relativity

The IIAG country scores and ranks are all relative, taking into account a country's performance in relation to the other 53 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.

Uncertainty

The results of the IIAG must be read alongside the standard errors, which measure the uncertainty around *Overall Governance* and category scores. Confidence intervals are published for the country scores for users to differentiate between country scores and changes over time. 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.