GLOBAL CORRUPTION INDEX 2018

BY GRP

Technical Methodology

Global Corruption Index 2018
Technical Methodology

The Global Corruption Index (GCI) is a composite index offering a classification of exogenous risks by country with regards to anti-corruption and anti-white collar crimes. The GCI was built in the framework of a more general risk mitigation plan, compliant with international recommendations and current binding legal requirements, such as the US FCPA, the UK Bribery act and the French law "Sapin II".

This technical methodology is meant to provide all necessary information for understanding the variables used and their respective impact in rankings.

1- Indicators and Data Overview

The GCI covers 199 countries and is composed of 28 variables constructed based on datasets that are exclusively borrowed from internationally recognized entities. Among these variables, 26 are for the calculation of a sub-index, the Corruption Index (CI), while the remaining 2 allow for an anti-white collar crimes (WCC) adjustment. As can be seen from tables 1 and 2 below, the GCI is divided into 6 indicators: (1) Ratification Status of Conventions, (2) Corruption Perception, (3) Corruption Experience, (4) Country Characteristics (5) Money Laundering and Terrorism financing and (6) Membership to FATF and / or related bodies.

Table 1 Levels of aggregation

	Ratification status of Conventions	Corruption Perception	Corruption Experience	Citizen's Voice and Transparency	Government Functioning and Effectiveness	Legal Context	Political Context	White Collar Crimes
Index	Global Corruption Index							
Weight	0.8							0.2
Sub-index	Corruption Index							
Weight	0.15	0.85						
Indicator		Corruption Evaluation						
Weight		0.3	0.2	0.5				
Indicator				Country Characteristics				
Weight				0.2	0.2	0.4	0.2	
Weight	0.9 / 0.1	0.5 / 0.25 / 0.25	0.35 / 0.35 / 0.3	0.31 / 0.23 / 0.23 / 0.23	0.25 / 0.2 / 0.15 / 0.15 / 0.15 / 0.1	0.5 / 0.1 / 0.1 / 0.1 / 0.1 / 0.1	0.7 / 0.3	0.9 / 0.1
Variable Ref.	V1, V2	V3, V4, V5	V6, V7, V8	V9, V10, V11, V12	V13, V14, V15, V16, V17, V18	V19, V20, V21, V22, V23, V24	V25, V26	V27, V28

The following table lists the data variables used to calculate the GCI, together with their respective variable reference code.

Table 2 Data Overview

Ref.	Variable
V1	Ratification status of the UN Convention against Corruption
V2	Ratification status of the OECD Anti-Bribery Convention
V3	Absence of Corruption
V4	Corruption Perception Index
V5	Control of Corruption
V6	Bribery Incidence
V7	Bribery Depth
V8	Bribery Rate
V9	Open Government
V10	Voice and Accountability
V11	Public Participation in the Budget Process
V12	Provision of Budget Information
V13	Rule of Law
V14	Regulatory Enforcement
V15	Government Effectiveness
V16	Budget Oversight
V17	Constraints on Government Powers
V18	Regulatory Quality
V19	Weakest Dimension of Justice
V20	Civil Justice is Free of Corruption
V21	Criminal System is Free of Corruption
V22	Criminal System is Impartial
V23	Civil Justice is Free of Improper Government Influence
V24	Criminal System is Free of Improper Government Influence
V25	Democracy Score
V26	Political Stability
V27	Money Laundering and Terrorism financing
V28	Members of the FATF and Related Bodies

A number of criteria were considered during the selection process:

- All data variables are linked to the measure of corruption and / or white collar crimes, either directly or indirectly. The direct measures of the Corruption Index consist of three surveys of perception and three surveys of reported experience, while the remaining variables are indirect measures meant to capture prevention mechanisms, related effects, causal effects and consequential effects. The objective of the latter group of measures is to unearth the latent information z (corruption) common to all selected variables

- In order to ensure cross-country comparability, no country specific information is considered. Such data would generate valuations relying on different bases / concepts, which is unsuitable for rankings

- Data sources with limited coverage are set aside. The lowest coverage rate considered is that of the variable V8, *Bribery Rate* (54% of country coverage)[1]

- Although some variables have wider coverage than others, none of them are limited to a specific cluster of countries. This decision is meant to guarantee that each variable offers a data set with scattered points across the full spectrum of possibilities

- During the selection process, preference is given to quantitative type of data. Qualitative information is also considered if and only if the transitivity axiom is ultimately satisfied

2- Missing Data

Several methods exist to deal with missing values, which can be grouped into two types of treatments: deletion - such as listwise deletion (complete-case analysis) and simple case deletion - or imputation.

Considering that most of our missing data is either of type MAR (Missing at Random) or MNAR (Missing not at Random), deletion is hardly appropriate and would lead to biased estimates. As the aim is to estimate corruption and white collar crimes, the lack of transparency for example is an important information. In this consideration, skipping countries for them to have a lack of due reporting would be counterproductive.

The processing of missing data is thus handled on a case-by-case basis depending on the structure of the datasets:

2.1- Imputation

First, in the case of time series datasets with visible trends, we proceed with a linear extrapolation from the five last available years. This method allows to estimating parameters based on real past values.

---

[1] The coverage rate considered is that obtained after the imputation processes based on real values, that is the linear extrapolation from the five available years and / or the LOCF, if applicable.

The second approach used is the method of the Last Observation Carried Forward (LOCF), which is a common statistical approach for time series data that consists in imputing the last available observation. Similar to the first method, only the last five available years are considered.

The two above-mentioned methods are selected to be based on known values that are specific to the countries, and consequently true at a point in time. In most cases, such methods can't be applied because no current nor past value is available. In these cases, we consider a third imputation method: Predictive Mean Matching (PMM) with multiple imputation.

Single imputation provides only one parameter estimate for each missing value and omits possible alternatives. It therefore tends to underestimate the standard errors and consequently overestimate the validity of the estimated scoring. As opposed to single imputation, multiple imputation provides *n* different possibilities for each missing value. These *n* possibilities allow for two desirable outputs:
-   First, each imputed value results from the pooling of the *n* parameter estimates, thus providing a better approximation of the true value
-   Second and more importantly, multiple imputation allows for measures of uncertainty, by sampling *n* times from the posterior predictive distribution.

As previously mentioned, the selected method of multiple imputations is that of Predictive Mean Matching (PMM). This approach allows us to preserve the distributions in the data and ensures that imputed values are plausible as it fills in values from real observations (Vink et al., 2014[2]). PMM provides a random value from a donor, based on the closeness of the regression-predicted values of the donor $\hat{\beta}$, with that of the recipient β^*. This implies that linear regressions are not used to generate imputed values but rather to determine the donor (Schenker, N. & Taylor, J.M.G., 1996[3]).

---

[2] Vink, G., Frank, L. E., Pannekoek, J., and van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. Statistica Neerlandica. 68(1). 61-90

[3] Schenker, N., & Taylor, J. M. G. (1996). Partially parametric techniques for multiple imputation. Computational Statistics & Data Analysis, 22(4), 425–446

The process by which PMM is performed is as follows (Vink et al., 2014[4]):

1. First, an Ordinary Least Squares (OLS) linear regression of γ given the selected predictors χ is performed to obtain the parameter estimates $\hat{\beta}$, $\hat{\sigma}^2$ and $\hat{\varepsilon}$, respectively the regression coefficient, the variance and the random error

2. In a second step, random draws of σ^{2*} and β^* are performed based on the posterior predictive distributions to provide new sets of coefficients. These draws allow for the calculation of $\hat{\gamma}_{missing}$

3. Predicted values are then generated by calculating $\hat{\gamma}$ for both cases with values (potential donors) and missing values (recipients), using the parameter estimates $\hat{\beta}$ and β^* respectively

4. The closeness of predicted values between donors and recipients is evaluated, so as to identify the three cases which minimizes $|\hat{\gamma}_{observed} - \hat{\gamma}_{missing}|$

5. Missing values are substituted from a random donor among those that satisfy the minimization criteria of the previous step.

6. Considering this index uses PMM for multiple imputation, the process starting from the random draws of σ^{2*} and β^* to the final imputation is repeated n times, in order to provide n complete datasets with n possible values for each missing case

2.2- Case Deletion

For some variables, no PMM imputation was performed and only true values were considered in the analysis. This is due to the structure of the data and the absence of correlation with other variables. In the case of a missing value, the algorithm proportionally redistributes the according weight to variables measuring the same indicator.

The variables with missing values are the following:

1. Bribery Rate
2. Bribery Incidence
3. Bribery Depth
4. Public Participation in the Budget Process
5. Budget Oversight

---

[4] Vink, G., Frank, L. E., Pannekoek, J., and van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. Statistica Neerlandica. 68(1). 61-90

The three first variables, namely *Bribery Rate*, *Bribery Incidence* and *Bribery Depth* represent the totality of the indicator *Corruption Experience*, which means that missing values in all three variables lead to an absence of *Corruption Experience* result for the concerned countries. Due to the explanation power of these three datasets, which report real experiences, these are still fully integrated in the analysis. Please find below the list of the 49 countries for which no *Corruption Experience* score is calculated. The corresponding weight is redistributed to the *Corruption Perception* indicator.

Countries with no data availability on corruption experience

Austria	Iran, Islamic Rep.	Oman
Bahrain	Ireland	Palau
Belgium	Kiribati	Qatar
Bermuda	Korea, Dem. Rep.	Saudi Arabia
Brunei Darussalam	Kuwait	Seychelles
Canada	Libya	Singapore
Cayman Islands	Liechtenstein	Somalia
Comoros	Luxembourg	Switzerland
Cook Islands	Macao SAR, China	Syrian Arab Republic
Cuba	Maldives	Turkmenistan
Denmark	Malta	Tuvalu
Equatorial Guinea	Marshall Islands	United Arab Emirates
Finland	Nauru	United Kingdom
France	Netherlands	United States
Greenland	New Zealand	Virgin Islands (U.S.)
Haiti	Niue	
Iceland	Norway	

3- Standardization

Aside from binary variables, all datasets were tested for skewness, then transformed and recoded if necessary. The mean and standard deviation is calculated and all variables are then standardized, to allow for a proper aggregation in the global scoring. Several normalization methods exist. The one used here is that of z-scores, which converts datasets to a common scale with a mean of zero and a standard deviation of one, as follows:

$$I_{i,c} = \frac{X_{i,c} - X_{i,\bar{c}}}{\sigma_{\bar{c}}}$$

with:
ι = variable
c = country
c̄ = reference country
σ = standard deviation

4- Aggregation

The aggregation process converts all data points to a scale of 0-100, where 0 represents the lowest risk of corruption and white collar crimes, and 100 corresponds to the highest risk of corruption and white collar crimes. Each country's global score is then calculated following the weights previously presented.

5- Measure of Uncertainty

Based on the *n* datasets obtained from the multiple imputation process, a standard error and a 90 percent confidence interval are calculated for each dataset to reflect the variance around the different scores.

Countries covered by the GCI Index 2018

1	Afghanistan	51	Dominica	101	Liberia	151	Saudi Arabia
2	Albania	52	Dominican Republic	102	Libya	152	Senegal
3	Algeria	53	Ecuador	103	Liechtenstein	153	Serbia
4	Angola	54	Egypt, Arab Rep.	104	Lithuania	154	Seychelles
5	Antigua and Barbuda	55	El Salvador	105	Luxembourg	155	Sierra Leone
6	Argentina	56	Equatorial Guinea	106	Macao SAR, China	156	Singapore
7	Armenia	57	Eritrea	107	Macedonia, FYR	157	Slovak Republic
8	Australia	58	Estonia	108	Madagascar	158	Slovenia
9	Austria	59	Ethiopia	109	Malawi	159	Solomon Islands
10	Azerbaijan	60	Fiji	110	Malaysia	160	Somalia
11	Bahamas, The	61	Finland	111	Maldives	161	South Africa
12	Bahrain	62	France	112	Mali	162	South Sudan
13	Bangladesh	63	Gabon	113	Malta	163	Spain
14	Barbados	64	Gambia, The	114	Marshall Islands	164	Sri Lanka
15	Belarus	65	Georgia	115	Mauritania	165	St. Kitts and Nevis
16	Belgium	66	Germany	116	Mauritius	166	St. Lucia
17	Belize	67	Ghana	117	Mexico	167	St. Vincent and the Grenadines
18	Benin	68	Greece	118	Micronesia, Fed. Sts.		
19	Bermuda	69	Greenland	119	Moldova	168	Sudan
20	Bhutan	70	Grenada	120	Mongolia	169	Suriname
21	Bolivia	71	Guatemala	121	Montenegro	170	Swaziland
22	Bosnia and Herzegovina	72	Guinea	122	Morocco	171	Sweden
23	Botswana	73	Guinea-Bissau	123	Mozambique	172	Switzerland
24	Brazil	74	Guyana	124	Myanmar	173	Syrian Arab Republic
25	Brunei Darussalam	75	Haiti	125	Namibia	174	Taiwan
26	Bulgaria	76	Honduras	126	Nauru	175	Tajikistan
27	Burkina Faso	77	Hong Kong SAR, China	127	Nepal	176	Tanzania
28	Burundi	78	Hungary	128	Netherlands	177	Thailand
29	Cabo Verde	79	Iceland	129	New Zealand	178	Timor-Leste
30	Cambodia	80	India	130	Nicaragua	179	Togo
31	Cameroon	81	Indonesia	131	Niger	180	Tonga
32	Canada	82	Iran, Islamic Rep.	132	Nigeria	181	Trinidad and Tobago
33	Cayman Islands	83	Iraq	133	Niue	182	Tunisia
34	Central African Republic	84	Ireland	134	Norway	183	Turkey
35	Chad	85	Israel	135	Oman	184	Turkmenistan
36	Chile	86	Italy	136	Pakistan	185	Tuvalu
37	China	87	Jamaica	137	Palau	186	Uganda
38	Colombia	88	Japan	138	Panama	187	Ukraine
39	Comoros	89	Jordan	139	Papua New Guinea	188	United Arab Emirates
40	Congo, Dem. Rep.	90	Kazakhstan	140	Paraguay	189	United Kingdom
41	Congo, Rep.	91	Kenya	141	Peru	190	United States
42	Cook Islands	92	Kiribati	142	Philippines	191	Uruguay
43	Costa Rica	93	Korea, Dem. Rep.	143	Poland	192	Uzbekistan
44	Cote d'Ivoire	94	Korea, Rep.	144	Portugal	193	Vanuatu
45	Croatia	95	Kuwait	145	Qatar	194	Venezuela, RB
46	Cuba	96	Kyrgyz Republic	146	Romania	195	Vietnam
47	Cyprus	97	Lao PDR	147	Russian Federation	196	Virgin Islands (U.S.)
48	Czech Republic	98	Latvia	148	Rwanda	197	Yemen, Rep.
49	Denmark	99	Lebanon	149	Samoa	198	Zambia
50	Djibouti	100	Lesotho	150	Sao Tome and Principe	199	Zimbabwe