ESG INDEX 2019

BY GRP

Technical Methodology

ESG Index 2019

Technical Methodology

The ESG Index (Environmental, Social and Governance Index) or ESGI is a composite measure offering a classification of exogenous risks by country with regards to the environment, human rights and health & safety. This index was built in the framework of a more general risk mitigation plan, compliant with international recommendations and current binding legal requirements, such as the French law "Devoir de Vigilance".

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 ESGI is based on 45 variables that are exclusively borrowed from internationally recognized entities. The ESGI is divided into 3 independent sub-indexes as presented below.

The Environment sub-index covers 180 countries and displays two indicators:

- (1) Ratification Status
- (2) Environmental Performance Index

The Human Rights sub-index covers 189 countries and comprises 4 indicators:

- (1) Ratification Status of Conventions
- (2) Social Rights
- (3) Civil and Political Rights
- (4) Collective Rights.

The last sub-index focuses on Health & Safety and provides a result for 184 countries. It is divided into 3 indicators:

- (1) Health
- (2) Safety
- (3) Inequality by Residential Area

177 countries and territories are scored in all three sub-indexes and are therefore included in the global scoring of the ESG Index

Table 1 Levels of aggregation

Index	Weight	Sub-index	Weight	Indicator	Weight	Indicator	Weight	Indicator	Weight	Variable Ref.
ESG Index	0.3	**Env.**	0.3	Ratification Status of the Kyoto Protocol and Paris Agreement					0.3	V1
									0.7	V2
			0.7	Environmental Performance Index					/	V3
	0.5	**Human Rights**	0.5	Ratified treaties, reporting compliance and standing invitations					0.55	V4
									0.15	V5
									0.15	V6
									0.15	V7
			0.35	Social Rights	0.5	Labor Rights			0.1	V8
									0.15	V9
									0.1	V10
									0.2	V11
									0.2	V12
									0.1	V13
									0.15	V14
					0.15	Education			0.7	V15
									0.3	V16
					0.15	Housing			0.5	V17
									0.5	V18
					0.2	Gender Equality			/	V19
			0.25	Civil and Political Rights	0.1	Public Affairs Investment			/	V20
					0.1	Press Freedom			/	V21
					0.3	Prohibition of Discrimination & Minority Rights	0.3	Factionalized Elites	/	V22
							0.3	Group Grievance	/	V23
							0.4	Perceptions	0.8	V24
									0.2	V25
					0.2	Personal Freedom			/	V26
					0.3	Political Rights and Civil Liberties Index			/	V27
			0.15	Collective Rights	0.7	Gobal Peace Index			/	V28
					0.3	Right to Self Determination			0.3	V29
									0.7	V30
	0.2	**Health & Safety**	0.4	Health Indicators	0.35	Medical Care			0.5	V31
									0.5	V32
					0.25	Life Expectancy			0.4	V33
									0.6	V34
					0.4	Access to Food and Drinking Water			0.65	V35
									0.35	V36
			0.4	Safety Indicators	0.5	Safety at Work			0.5	V37
									0.5	V38
					0.5	Social Security	0.4	Social health Protection Coverage	/	V39
							0.3	Unemployment Benefits	/	V40
							0.3	Protection of Vulnerable Persons	0.5	V41
									0.5	V42
			0.2	Inequality by Residential Area					0.4	V43
									0.3	V44
									0.3	V45

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

Table 2 Data Overview

[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]
	[illegible]	[illegible]

A number of criteria were considered during the selection process:

- All data variables are linked to the measures of environment, human rights and health & safety, either directly or indirectly.

- 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, except in the case of a high explanation power. The two variables displaying the highest number of missing values are V37 *Non-fatal Occupational Injuries* (34.2% of country coverage) and V38 *Fatal Occupational Injuries* (35.3% of country coverage)[1]. With the objective to measure health & safety risks, injuries at work are of primary importance. In this context and considering the absence of other similar measures, both variables were kept in the analysis.

- 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.

[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 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.

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

[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

linear regressions are not used to generate imputed values but rather to determine the donor (Schenker, N. & Taylor, J.M.G., 1996[3]).

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. Income Inequality
2. Migrant Acceptance Index

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

[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

3. Freedom to Make Life Choices
4. Non-fatal Occupational injuries
5. Fatal Occupational injuries
6. Mandatory Paid Maternity Leave
7. Inequality by Residential Area of Health Care Deficit

Despite the process described above, some missing values remained. A redistribution of weights is thus also performed between indicators. The concerned indicators are the following:

1. Perception (Prohibition of Discrimination & Minority Rights)
2. Public Affairs Investment
3. Safety at Work

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

$\bar{c}$ = 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, and 100 corresponds to the highest risk. Country scores are then calculated for each sub-index (Environment, Human Rights and Health & Safety) using an arithmetic mean and following the weights previously presented (table 1).

In order to provide a unique risk score encompassing all three measures, scores by sub-index are eventually aggregated using a weighted geometric mean.

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.

Global Risk Profile Sàrl
91 Rue de Lyon
1203 Geneva
Switzerland

Global Risk Profile Sàrl
Email:info@risk-indexes.com
CH +41 (0) 22 794 39 11
FR +33 (0) 1 73 19 11 72
Web. www.globalriskprofile.com

Global Risk Profile France
27 Avenue de l'Opéra
75001 Paris
France

Index Research & Development
Sonia THURNHERR
Lead Data Scientist

