# Contents

## 1 Overview

## 2 The underlying model

### 2.1 The data hierarchy

### 2.2 Weighting factors, dimensions and sources

### 2.3 Banding and uncertainty

## 3 The system and database

### 3.1 Data transformations

#### 3.1.1 Base transformations

#### 3.1.2 High-value/high-index and high-value/low-index

#### 3.1.3 Standardisation

### 3.2 Master index and factor index calculation

### 3.3 Banding calculation

## 4 Glossary of terms

## List of figures

### 1 The country risk ranking hierarchy
1 OVERVIEW

COUNTRY RISK RANKING is a system built from the model outlined below that aims to provide an analysis of countries based on existing online sources of information about those countries. The intention is to combine relevant information from different sources, by performing a range of calculations, in a manner that can be analyzed and investigated, to provide what we call the Master Index. The system is intended to help the user interrogate the sources, calculations and results, from a single location, in an intuitive and statistically meaningful way.

The intention of COUNTRY RISK RANKING is to instantiate the Master Index as an indicator of operational risk for individual countries, though we omit consideration of the semantics of the notion of risk, and its relation to the COUNTRY RISK RANKING Master Index. Thus COUNTRY RISK RANKING does not provide an assessment of risk per se, but provides an aggregation of different factors to derive an index that might in turn be used, subjectively, as an indication of the relative standing of countries with respect to the information identified. In other words, COUNTRY RISK RANKING should be viewed as summary measure of a range of constituent information that might be considered to be associated with operational risk.

More specifically, the model underlying COUNTRY RISK RANKING allows data from a range of different sources for a set of countries to be grouped together in a statistically meaningful way so that an overall Master Index value (between 0 and 10) for each country can be ascertained.

Users may simply employ the COUNTRY RISK RANKING system with standard supplied settings (known as weights). Alternatively, users may tailor the COUNTRY RISK RANKING system to influence both the source information that is used and the importance and relevance that is attached to that information. Finally, users may also interrogate the COUNTRY RISK RANKING Master Index to analyze various aspects of the source information that derives the index.

COUNTRY RISK RANKING is designed to be as transparent as possible in its calculations to derive the resulting index, to facilitate tailoring of the system to individual requirements, and to provide the requisite understanding for users to be able to determine whether and how to do so.
2 THE UNDERLYING MODEL

2.1 The Data Hierarchy
The raw data for COUNTRY RISK RANKING originates from information freely available on the Web (and which may be obtained directly from the COUNTRY RISK RANKING system, with each distinct data set being referred to as a source. These feed into COUNTRY RISK RANKING dimensions, representing a particular identified area of interest, such as freedom of the press or military expenditure. Thus, a dimension contains one or many sources, but a source is only ever associated with a single dimension. The dimensions that group sources together are, in turn, grouped into factors, which provide a third level in the information hierarchy, and it is these that are finally combined to generate the COUNTRY RISK RANKING Master Index. The relationship of sources, dimensions and factors is shown in Figure 1.

There are three factors in the current model, political, financial and criminal. Since the source data is varied and inconsistent across different sources, it must be transformed into a standard form before it can be combined with other source data to feed into dimensions, which constitute the second level of the hierarchy. As indicated above, correlated dimensions are grouped together to derive factors, which constitute the third level of the hierarchy. Finally, these factors are combined at the top level, before undergoing a transformation to generate a Master Index between 0 and 10, which places countries in rank order and may be used, in particular, to compare the status of different countries.

In order to generate the Master Index from the disparate sources via dimensions and factors in this way, we make several underlying assumptions about the nature of the data.

2.2 Weighting Factors, Dimensions and Sources
As we have stated earlier in this document, once the data for each source is transformed into some appropriate representation as detailed above, it is then combined to derive the Master Index. There are three stages to this.

1. First, for a particular dimension, all the sources are aggregated together as a weighted mean. For each source, a weight is provided that biases the mean according to the “importance” of the data. The sources weights are specified by World-Check as defaults reflecting expert user opinion. We call these weights - source weights. Changing the Dimension weights can be set by the user to reflect the “relevance” and “accuracy” of the source grouping.

2. Then, for each factor, dimensions are combined in a similar way, with each dimension receiving a user-specified weight before combination. However, here we need to be careful to ensure that the dimension weights are meaningful when considered in relation to the sources. For example, if Dimension A has only one source, which is weighted very low, reflecting poor reliability, then it doesn’t make sense to weight Dimension A very highly. To address this, we may need to adjust downward the weight of Dimension A in response to its low source weight. Of course in other situations is possible that the value of the system dimension weight is tuned upwards from the value of the user dimension weight. At this level, therefore, we use two weights. First, we use the user-specified weight, called a user dimension weight, and indicating its relevance, as specified by the user, in determining its relative contribution to a factor. However, we introduce another weight, derived through statistical methods from the...
source weights and the user dimension weight to give an adjusted system dimension weight that is used in the basic weighted mean calculation. The system dimension weight for a given dimension can vary across countries, as different countries may have different sources available.

3. Finally, we can weight factors as determined by the user. These are used in a simple weighted mean calculation to derive the final index.

If a user wishes to “switch off” a parameter (dimension or factor) the weighting should be set to 0. Thus, for example, to obtain a COUNTRY- CHECK Master Index built out of only one factor, the user will need to set the weights of the other factors to 0. In such a case, of course, there would be no distinction between the index of the factor and the COUNTRY RISK RANKING Master Index and the system would thus give a value of between 0 and 10 for an individual factor.

2.3 Banding and Uncertainty
The model also allows uncertainty to be quantified. Firstly, the master index scale is mapped to a number of bands. For each country the probability, based on the available data, of band membership for each band is computed. The most probable band is returned, together with a list of subsidiary bands (those with relatively high probability compared with the highest probability band). The full list of band probabilities can also be returned. The system has the flexibility to allow users to determine (within limits) the number of bands, and the proportion of countries expected (under the model) to be returned in the extreme bands.
3 THE SYSTEM AND DATABASE

3.1 Data Transformations

Since some raw data can be very "skewed" in the sense that it can range over many very small values and few very high values, it must be transformed (scaled appropriately) to facilitate easy comparison between different source data. This enables us to ensure consistency with the underlying statistical model and algorithmic method, while at the same time providing meaningful results.

Similarly, some raw data simply provides values on an ordinal scale in the sense that only a few numbers are actually used (for example, integers in the range 1–7), and this must also be mapped onto an appropriate scale.

3.1.1 Base Transformations

We have therefore developed a set of transformations for the various sources so that they can be used to derive an index. In total, there are three transformations that can cover the different sources and can be applied to the source data, as follows.

1. Log transformation

When there is non-negative data with a skewed distribution, we use a logarithmic transformation (as is common in statistical analysis) to make the distribution more symmetric, and avoid the extreme positive values having an inappropriate impact on our analysis.

Occasionally zeros are present in the data, for which the logarithmic transformation is inappropriate. In such cases, the value of zero is replaced by a representative value, consistent with the underlying statistical model.

2. Categorical (including binary)

Where the data fall in one of two or more ordered categories, but the numerical values assigned to those categories are essentially arbitrary, then it is possible to provide numerical labels so that the underlying distribution more closely reflects the statistical model.

3. Uniform

In a very few cases, the source data are more uniformly distributed across their range than would be expected under the model. In such cases, a transformation is used to ‘normalize’ the values.

3.1.2 High-Value/High-Index and High-Value/Low-Index

It is important to note that the model makes no assumptions about whether a high score should contribute to a high index. Therefore the final step of the transformation is to specify this relationship (1 or -1) and the data is multiplied by this number. Here, +1 indicates that a high value gives a high index, while -1 indicates that a low value gives a high index.

In this context, at the end of the transformation, a high score contributes to a high index.

3.1.3 Standardization

A further transformation is then applied to each of the transformed source data sets. We call the resulting data, standardized data which has a mean of 0 and variance consistent with the underlying model assumptions.

Note that while transformed data is fixed for each source, the standardized data changes whenever user weights are changed for either source or dimension.

3.2 Master Index and Factor Index Calculation

- The value of a dimension is the mean of the standardized values of the constituent sources according to the source weights.
- A factor index is the mean of the dimension values weighted according to the system dimension weights.
- The aggregate index is an intermediate index that is the weighted mean of the factor indices, with weights determined by the user.
- The Master Index is then the result of transforming the aggregate index to place it appropriately on a scale of between 0 and 10.

3.3 Banding Calculation

- The masterindex scale is mapped to a number of bands.
- For each country the probability based on the available data, of band membership for each band is computed.
- The most probable band and subsidiary bands are identified.
4 GLOSSARY OF TERMS

- **Master Index.** The overall score, between 0 and 10, gained by a country.
- **Aggregate Index.** The weighted mean of the factor indices.
- **Factor Index.** This is determined from the weighted mean of the constituent dimension values. Note that to determine a factor index on the same scale as the Master Index in this model, the weights of the other factors should be set to 0.
- **Dimension Value.** The weighted mean of the standardized values of the constituent sources.
- **Source.** The original online data for a single measure.
- **Source Provider.** The hosting organization or provider for the source.
- **Source Weight.** The user's measure of both the accuracy of the source and its relevance with respect to the associated dimension.
- **User Dimension Weight.** The user's measure of the relevance of a dimension with respect to a factor.
- **System Dimension Weight.** Produced by the system from the user dimension weight and associated source weights. It is this value that is used in the basic weighted mean calculation.
- **Factor Weight.** The user's measure of the relevance of a factor with respect to the overall Master Index.
- **Source data.** The values contained in COUNTRY RISK RANKING data files resulting from data gathering from sources.
- **Transformed data.** The data after one of several available transformations takes the source data closer to a normal distribution. In addition, transformed data is such that a high value contributes to a high index.
- **Standardized data.** The data that results from taking the transformed data and placing it on the same scale.
- **Banding.** An ordinal discretization of the master index scale.

THOMSON REUTERS

The Thomson Reuters Governance, Risk & Compliance (GRC) business delivers a comprehensive set of solutions designed to empower audit, risk and compliance professionals, business leaders, and the Boards they serve to reliably achieve business objectives, address uncertainty, and act with integrity.

Thomson Reuters connects business transactions, strategy and operations to the ever-changing regulatory environment, enabling firms to manage business risk. A comprehensive platform supported by a range of applications and trusted regulatory and risk intelligence data, brings together market-leading solutions for governance, risk and compliance management, global regulatory intelligence, financial crime, anti-bribery and corruption, enhanced due diligence, training and e-learning, and board of director and disclosure services.

Visit risk.thomsonreuters.com

For more information, contact your representative or visit us online.