

A short methodological note

Transparency International Corruption Perceptions Index (CPI) 2006

1. The CPI gathers data from sources that span the last two years (for the CPI 2006, this includes surveys from 2005 and 2006).
2. All sources provide a ranking of countries, i.e., include an assessment of multiple countries.
3. All sources measure the overall extent of corruption (frequency and/or size of bribes) in the public and political sectors.
4. For CPI sources that are surveys, and where multiple years of the same survey are available, all annual data are included to provide a smoothing effect.
5. For sources that are scores provided by experts (risk agencies/country analysts), only the most recent iteration of the assessment is included, as these scores are generally peer reviewed and change very little from year to year.
6. Evaluation of the extent of corruption in countries is done by country experts, non resident and residents (in the CPI 2006, this consists of the following sources: CPIA, EIU, FH, MIG, UNECA and WMRC); and resident business leaders evaluating their own country (in the CPI 2006, this consists of the following sources: IMD, PERC, and WEF).
7. To determine the mean value for a country, standardisation is carried out via a matching percentiles technique. This uses the ranks of countries reported by each individual source. This method is useful for combining sources that have a different distribution. While there is some information loss in this technique, it allows all reported scores to remain within the bounds of the CPI, that is to say, to remain between 0 and 10.
8. A beta-transformation is then performed on scores. This increases the standard deviation among all countries included in the CPI and avoids the process by which the matching percentiles technique results in a smaller standard deviation from year to year.
9. Next, all values for a country are averaged, to determine a country's score.
10. The CPI score and rank are accompanied by the number of sources, high-low range, standard deviation and confidence range for each country.
11. The confidence range is determined by a bootstrap (non-parametric) methodology, which allows inferences to be drawn on the underlying precision of the results. A 90% confidence range is then established, where there is 5% probability that the value is below and 5% probability that the value is above this confidence range.
12. Research shows that the unbiased coverage probability for the confidence range is lower than its nominal value of 90%. The accuracy of the confidence interval estimates increases with a growing number of sources: for 3 sources, 65.3%; for 4 sources, 73.6%; for 5 sources, 78.4%; for 6 sources, 80.2%; and for 7 sources, 81.8%.
13. The overall reliability of data is demonstrated in the high correlation between sources. In this regard, Pearson's and Kendall's rank correlations have been performed, which provided average results of .87 and .72 respectively.