European Commission

JRC SCIENTIFIC AND POLICY REPORTS

Corruption Perceptions Index 2012 Statistical Assessment

Michaela Saisana, Andrea Saltelli

Report EUR 25623 EN

Joint Research Centre

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Michaela Saisana
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 361, 21027 Ispra (VA), Italy
E-mail: michaela.saisana@jrc.ec.europa.eu
Tel.: +39 0332 78 6572
Fax: +39 0332 78 5733

http://composite-indicators.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 77239

EUR 25623 EN

ISBN 978-92-79-27732-0

ISSN 1831-9424

doi: 10.2788/69609

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

Table of Contents

List of Tables

List of Figures

Executive Summary

The Corruption Perceptions Index (CPI) has been developed since 1995 by Transparency International as a composite indicator that measures perceptions of corruption in the public sector in different countries around the world. It does so by aggregating different sources of corruption-related data that are produced by a variety of independent and well known institutions, such as the World Bank, the World Justice Project, the African Development Bank, the Economist Intelligence Unit and other.

The European Commission Joint Research Centre (JRC) in Ispra-Italy was invited by the Transparency International to assess the new methodology used to develop the CPI 2012 and to shed more light into the consequences that come with this change. The JRC analysis was based on in-house quality control process that aims to ensure the transparency of the methodology and the reliability of the results. The statistical assessment of the CPI 2012 was done along three main avenues: an evaluation of conceptual/statistical coherence of the index structure, an interpretation of the rankings based on significance tests, and an evaluation of the impact of key modelling assumptions (imputation and normalisation) on countries' scores and ranks.

The report discusses why the old methodology does not fully achieve one of the main principles behind the development of the CPI, which is to "compensate" for eventual errors among sources and to provide a more reliable picture of the perceived level of corruption around the world than would any of the thirteen sources taken independently.

The statistical coherence of the 2012 CPI is based on an analysis of the covariance structure across the thirteen sources of information. It shows that the high correlation between the CPI ranking and the sources is not a symptom of redundancy but is driven by the fact that all sources attempt to measure the same phenomenon, which is the perceived level of corruption in the public sector. The analysis also provides a statistical justification on the use of simple average across the

sources. Multiple comparison tests after Bonferroni correction suggest that there seems to be no bias in the CPI scores with respect to the number of sources used, whilst countries with few available sources tend to have slightly larger standard errors (on average) compared to countries that are evaluated using more sources. Nevertheless, the criterion for a country's inclusion in the CPI if evaluated by at least three sources seems to be sufficient. A recommendation is made on the calculation of the standard errors, which are currently overestimated by the current formula.

The modeling assumptions (normalization coupled with estimation of missing data) are found to have a moderate impact on the CPI ranking (no impact for 52 countries, less than five-rank shift for 94% of the countries). The analysis also shows that the maximum shift with respect to the CPI rank when excluding one of the sources is up to 4 positions for 75% of the countries, which suggests that no source dominates the CPI and that all sources contribute to determining the CPI ranking in a balanced way.

Altogether, the statistical analyses described in this report underline the contribution of the CPI to the measurement of perceived corruption in the public sector at national level worldwide:

- the CPI covers more countries than any of the individual sources alone,
- the CPI may be more reliable than each source taken separately,
- the CPI can efficiently differentiate the level of corruption between countries, unlike some sources where a large number of countries is assessed at the same level of corruption,
- the CPI reconciles different view points on the issue of corruption, noteworthy since no country is classified as better off than another country on all common sources.

The main recommendation for the CPI team is to adjust the formula for the standard errors for the small population size (errors that are currently overestimated) and for policy makers to consider the statistical significance (by means of effect size for example) when comparing the CPI scores. The results make clear that even when differences in the CPI country scores are statistically significant they should be carefully interpreted.

1. Introduction

The Corruption Perceptions Index (CPI) has been developed since 1995 by Transparency International as a composite indicator that measures perceptions of corruption in the public sector in different countries around the world. It does so by aggregating different sources of corruption-related data that are produced by a variety of independent and well known institutions. During the past 17 years, the CPI has evolved as both the sources used to compile the index and the methodology have been adjusted and refined.

Combining different sources of corruption-related data that come from the World Bank, World Justice Project, African Development Bank, Economist Intelligence Unit and other, as done in the CPI, is both advantageous but also potentially worrisome. The main advantage and *added value* of the CPI lays in the fact that an index that aggregates a set of independent sources that measure the same perceived concept can be more reliable than each source taken separately. It also raises practical challenges related to the quality of available data and the combination of these into a single number.

The European Commission Joint Research Centre (JRC) in Ispra-Italy was invited by the Transparency International to assess the new methodology used to develop the CPI 2012 and to shed more light into the consequences that come with this change. The JRC has researched extensively on the complexity of composite indicators and ranking systems that classify countries' performances along policy lines (Saisana *et al.*, 2005; 2011; Saltelli *et al.* 2008). The JRC analyzed the revised methodology of the 2012 CPI based on in-house[1] quality control process in order to ensure the transparency of the methodology and the reliability of the results. This should enable policymakers to derive more accurate and meaningful conclusions.

The statistical assessment of the CPI 2012 was done along three main avenues: an evaluation of conceptual/statistical coherence of the index structure, an interpretation of the rankings based on significance tests, and an evaluation of the impact of key modelling assumptions (imputation and normalisation) on countries' scores and ranks.

[1] The JRC analysis was based on the recommendations of the OECD (2008) Handbook on Composite Indicators, and on more recent research from the JRC implemented in numerous auditing studies of composite indicators available at http://composite-indicators.jrc.ec.europa.eu/

The report is structured as follows.

Section 2 presents the thirteen sources that were used in the 2012 CPI, the revised methodology used in the 2012 CPI and the old methodology based on ranks that had been used in past releases of the index. It discusses why the old methodology does not fully achieve one of the main principles behind the development of the CPI, which is to "compensate" for eventual errors among sources and to provide a more reliable picture of the perceived level of corruption around the world than would any of the thirteen sources taken independently.

Section 3 analyzes the statistical coherence of the 2012 CPI based on an analysis of the covariance structure across the thirteen sources of information. It shows that the high correlation between the CPI ranking and the sources is not a symptom of redundancy but is driven by the fact that all sources attempt to measure the same phenomenon, which is the perceived level of corruption in the public sector. The analysis described herein also provides a statistical justification on the use of simple average across the sources. Multiple comparison tests after Bonferroni correction suggest that there seems to be no bias in the CPI scores with respect to the number of sources used, whilst countries with few available sources tend to have slightly larger standard errors (on average) compared to countries that are evaluated using more sources. Nevertheless, the criterion for a country's inclusion in the CPI if evaluated by at least three sources seems to be sufficient. A recommendation is made on the calculation of the standard errors, which are currently overestimated by the current formula.

Section 4 discusses how to interpret the difference between two countries scores by employing Cohen's effect size. Overall, the CPI ranking accurately reflects when country differences are significant or not. A suggestion for policy makers is that even significant differences should be carefully interpreted given that there might be a substantial overlap in the resulting distributions for the countries.

Section 5 assesses the impact of modeling assumptions (normalization coupled with estimation of missing data) on the CPI ranking, and it is found that there is absolutely no difference between the CPI ranking and the simulated ranking for 52 countries, whilst there is less than five-rank difference for 94% of the countries. The analysis also shows that the maximum shift with respect to the CPI rank when excluding one of the sources

is up to 4 positions for 75% of the countries. This suggests that no source dominates the CPI and that all sources contribute to determining the CPI ranking in a balanced way.

Section 6 concludes.

2. CPI- Sources and methodology

The measurement of the perceived level of corruption by Transparency International is an evolving project since 1995. Every year, such measurement builds upon previous editions while refined with newly available data. The 2012 CPI is calculated for 176 countries around the world. The thirteen sources of information used to build the CPI are listed in Table 1. The sources differ in the number of countries covered, ranging from 16 countries covered in the Political and Economic Risk Consultancy Asian Intelligence to 175 countries covered in the Global Insight Country Risk Ratings. More detailed information on the sources and the rationale for inclusion of each source is offered in the main report of the 2012 CPI.

Table 1. 2012 CPI Sources of Information

Source	Number of countries
1. African Development Bank Governance Ratings (AFDB)	53
2. Bertelsmann Foundation Sustainable Governance Indicators (BF-SGI)	31
3. Bertelsmann Foundation Transformation Index (BF-BTI)	128
4. Economist Intelligence Unit Country Risk Ratings (EIU)	138
5. Freedom House Nations in Transit (FH)	29
6. Global Insight Country Risk Ratings (GI)	175
7. IMD World Competitiveness Yearbook (IMD)	59
8. Political and Economic Risk Consultancy Asian Intelligence (PERC)	16
9. Political Risk Services International Country Risk Guide (ICRG)	140
10. Transparency International Bribe Payers Survey (TI)	29
11. World Bank - Country Performance and Institutional Assessment (WB)	67
12. World Economic Forum Executive Opinion Survey (WEF)	147
13. World Justice Project Rule of Law Index (WJP)	97

Source: Corruption Perceptions Index 2012

The most recently released country scores from those thirteen sources were used in the development of the CPI 2012. Countries were included if they were evaluated by at least three sources; this was the case for 19 countries (e.g. Barbados, Bhutan, North Korea). The maximum number of sources based on which a country was evaluated was ten; this was the case for six countries – Poland, South Korea, Hungary, Czech Republic and India. Most countries were evaluated using 7-8 sources.

The CPI is refined every year in a transparent exercise to improve the way perceived corruption in the public sector is measured.

Old methodology

In past releases of the index, the normalization method that was used to place the sources into a common scale was a matching percentiles technique. That approach considered country ranks on each source. It was useful for combining sources that had different distributions. It also allowed all reported scores to remain within the bounds of the CPI ([0, 10]). A beta-transformation was then applied to the normalized scores. This increased the standard deviation among all countries included in the CPI and made it possible to differentiate more precisely between countries that appeared to have similar scores. A main limitation of that approach is the information loss due to the fact that only country ranks are considered and not the relative distance between them in a given source of information. This does not fully achieve one of the main principles behind the development of the CPI, which is to "compensate" for eventual errors among sources by taking their average. The notion of compensation is strongly linked to relative distances as opposed to ranks.

New methodology

Upon these conceptual considerations and also for simplicity in communication and to allow comparisons over time, the CPI 2012 is calculated using a simple average of standardized scores. More specifically, all thirteen sources are standardized by subtracting the mean of the data and dividing by the standard deviation (z-scores) and then rescaled to have a mean 45 and standard deviation 20.

The formula for the standardization is: $= \frac{x_i - mean(x)}{std(x)} \times sign \times 20 + 45$

The direction of the effect of the source is taken into account at this stage. For sources, for which the lower the value of the source, the less the perceived level of corruption, a negative sign is used. This is done for five sources: Economist Intelligence Unit Country Risk Ratings, Freedom House Nations in Transit, Global Insight Country Risk Ratings, Political and Economic Risk Consultancy Asian Intelligence, and Transparency International Bribe Payers Survey.

After the standardization, any values beyond the 0-100 scale are capped. For the normalized scores to be comparable between the thirteen sources, the mean and standard deviation need to be defined as global parameters. In other words, what would the mean and standard deviation of each source would have been if all 176 countries had been evaluated by each source? To this end, the CPI 2012 uses the "impute" command in the statistical software package STATA in order to impute scores for all those countries that are missing data in each source. The mean and standard deviation for each source across the 176 countries are then calculated and used as the parameters to standardize the sources during the normalization. An important remark is that the imputed values are used only during the calculation of the 'global mean and standard deviation' but not for the calculation of CPI country scores, which are subsequently calculated as simple averages of the normalized scores across the available sources only. The CPI scores are in the range 0 to 100 (=lowest level of perceived corruption).

3. Conceptual and statistical coherence in the CPI

Each of the thirteen sources included in the CPI measures the overall extent of corruption (frequency and/or size of corrupt transactions) in the public and political sectors and provides a ranking of countries that reflects the "perception of corruption" in the countries covered by each source. The aim of the CPI is to provide a more reliable picture of the perceived level of corruption around the world than would any of the thirteen sources taken independently.

Assessing potential redundancy of information in the CPI

The country rankings from the thirteen different sources tend to correlate well with each other. There is also a high correlation between the CPI ranking and each of the sources (Table 3). These high correlations were expected, given that all sources attempt to measure the same phenomenon, that is the perceived level of corruption in the public sector. Despite the high correlations among the CPI sources, the information offered by the CPI is not redundant. In fact, the thirteen sources cover different countries− from 16 countries for the Political and Economic Risk Consultancy Asian Intelligence to 175 countries for the Global Insight Country Risk Ratings. Hence, combining the

information on the perceived level of corruption from these different sources, as done in the CPI, brings the advantage of covering more countries than any of the individual sources alone, while at the same time may be more reliable than each source taken separately. In the CPI 2012, there is one more country/territory included − Kosovo− besides the 175 countries classified in the Global Insight Country Risk Ratings. Furthermore, the CPI can efficiently differentiate the level of corruption between countries, unlike some sources where a large number of countries is assessed to have the same perceived level of corruption (e.g. 40 in the Global Insight Country Risk Ratings). One more feature of the CPI is that it reconciles different view points on the issue of corruption. If the countries' classifications in the thirteen sources were to be taken at face value, it is found that no country is classified as better off than another country on all common sources. This is an important remark which adds to the contribution of the CPI in the measurement of perceived corruption at national level worldwide.

Principal Component Analysis was applied to the six sources with the widest country coverage, namely BF-BTI, ICRG, WEF, WJP, EIU, GI (67 countries are common to all sources).[2] The first latent dimension accounts for 78% of the total variability in the six sources (see Table 2). Furthermore, the six sources have nearly equal weights and loadings[3] on the first latent dimension. These results suggest that assuming equal weights and an arithmetic average to aggregate the six sources is statistically supported by the data. In more practical terms however, equal weights in the case of the CPI may be justified on the premise that all these sources are very important and that there is no a priori rationale for giving a higher weight to one source than to another.

Table 2. Principal Component Analysis on six CPI sources

PC	Eigenvalue	Variance explained (% total)	Source	Loadings on the first PC
1	4.70	78.3	BF-BTI	0.90
2	0.42	85.3	ICRG	0.84
3	0.37	91.4	WEF	0.83
4	0.20	94.8	WJP	0.91
5	0.19	97.8	EIU	0.89
6	0.13	100.0	GI	0.93

Source: Saisana and Saltelli, 2012, European Commission Joint Research Centre

[2] PCA could not be applied to the entire set of 13 sources as they do not have any single country in common.

[3] A loading in principal component analysis is the correlation coefficient between a variable and the Principal Component (latent dimension).

Table 3. Spearman rank correlations and Gamma statistics for the CPI sources

	CPI	AFDB	BF (SGI)	BF (BTI)	IMD	ICRG	WB	WEF	WJP	EIU	GI	PERC	TI	FH
CPI		0.71	0.86	0.78	0.80	0.84	0.82	0.71	0.77	0.93	0.90	0.90	0.67	0.84
AFDB	0.79 (n=53)			0.64		0.62	0.78	0.33	0.39	0.59	0.49			
BF (SGI)	0.92 (n=31)			0.79	0.75	0.86		0.76	0.70	0.89	0.88			
BF (BTI)	0.88 (n=128)	0.66 (n=43)	0.73 (n=8)		0.42	0.60	0.73	0.40	0.50	0.82	0.75	0.73	0.51	0.90
IMD	0.93 (n=59)	(n=1)	0.86 (n=31)	0.54 (n=34)		0.75		0.80	0.73	0.80	0.83	0.79	0.52	0.67
ICRG	0.87 (n=140)	0.6 (n=37)	0.87 (n=31)	0.61 (n=105)	0.85 (n=59)			0.64	0.71	0.83	0.84	0.70	0.69	0.74
WB	0.83 (n=67)	0.73 (n=37)	0 (n=0)	0.72 (n=56)	0 (n=1)	(n=40)		0.31	0.39	0.84	0.63			0.73
WEF	0.86 (n=147)	0.39 (n=40)	0.86 (n=31)	0.52 (n=108)	0.94 (n=59)	0.72 (n=126)	0.34 (n=48)		0.62	0.71	0.68	0.76	0.62	0.41
WJP	0.91 (n=97)	0.42 (n=21)	0.8 (n=26)	0.63 (n=77)	0.89 (n=50)	0.78 (n=90)	0.44 (n=30)	0.78 (n=95)		0.79	0.74	0.89	0.75	0.64
EIU	0.92 (n=138)	0.48 (n=33)	0.85 (n=31)	0.76 (n=100)	0.85 (n=59)	0.78 (n=124)	0.67 (n=36)	0.76 (n=126)	0.8 (n=91)		0.94	0.96	0.58	0.81
GI	0.94 (n=175)	0.48 (n=53)	0.84 (n=31)	0.78 (n=127)	0.9 (n=59)	0.83 (n=140)	0.57 (n=66)	0.77 (n=147)	0.82 (n=97)	0.9 (n=138)		0.87	0.68	0.94
PERC	0.97 (n=15)	(n=0)	(n=4)	0.8 (n=11)	0.91 (n=13)	0.79 (n=14)	(n=3)	0.88 (n=15)	0.96 (n=14)	0.96 (n=15)	0.91 (n=15)		0.60	
TI	0.83 (n=29)	(n=5)	(n=13)	0.63 (n=22)	0.7 (n=24)	0.78 (n=29)	(n=5)	0.78 (n=29)	0.89 (n=29)	0.68 (n=29)	0.78 (n=29)	0.76 (n=10)		
FH	0.94 (n=29)	(n=0)	(n=4)	0.95 (n=29)	0.74 (n=13)	0.73 (n=20)	0.76 (n=8)	0.5 (n=25)	0.78 (n=20)	0.82 (n=21)	0.94 (n=28)	0 (n=0)	(n=4)	

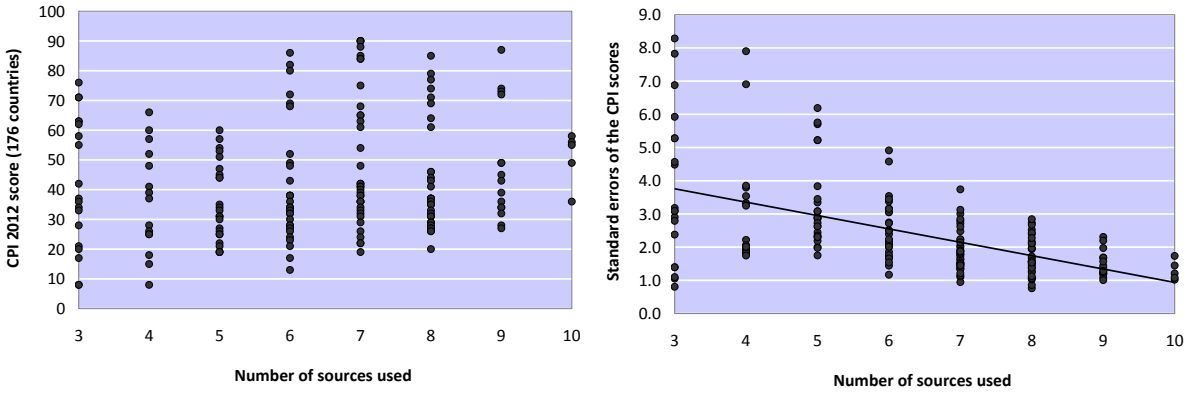
Source: Saisana and Saltelli, 2012, European Commission Joint Research Centre

Notes: Low diagonal: Spearman rank correlation coefficients (significant at 5% level). Number of countries that are common to each pair of sources is given in the parenthesis. Upper diagonal: Gamma statistic (significant at the 5% level), which is to be preferred over the Spearman rank correlation for sources with tied values, namely AFDB, WB, GI, BF-TI, FH, BF-SGI, EIU. All coefficients are positive because sources where lower scores represent lower levels of corruption were reversed by multiplying every score in the data by -1.

Assessing potential bias introduced in the CPI

A legitimate question is whether the CPI scores or the standard errors associated with them are biased with respect to the number of sources that were used to evaluate each country (ranging from three sources that were used to evaluate 19 countries, up to 10 sources that were used to evaluate five countries, see Figure 1). A multiple comparison test after Bonferroni correction[4] was used for the comparison of the means of the CPI country scores grouped per number of sources. The results suggest that there is no pattern between the CPI score and the number of sources that were used to evaluate a country. In fact, the eight group means of the CPI scores for 3, 4, up to 10 sources, are not different from each other at the 5% level. Hence, the CPI scores are not biased to the number of sources that were used to evaluate each country.

Figure 1. Impact of number of sources on the CPI scores and standard errors



Source: Saisana and Saltelli, 2012, European Commission Joint Research Centre

Before discussing whether there is a pattern between the standard errors associated to the CPI scores and the number of sources used to evaluate each country, we should add an important remark on the calculation of the standard error of the mean, which often goes unnoticed in the relevant literature. The standard error of the mean is often calculated as the ratio of the standard deviation over the square root of the sample size:

[4] When performing a simple t-test of one group mean against another, one needs to specify a significance level that determines the cutoff value of the t-statistic. For example, one can specify the value alpha = 0.05 to insure that when there is no real difference, one will incorrectly find a significant difference no more than 5% of the time. When there are many group means, there are also many pairs to compare. If one applied an ordinary t-test in this situation, the alpha value would apply to each comparison, so the chance of incorrectly finding a significant difference would increase with the number of comparisons. Multiple comparison procedures are designed to provide an upper bound on the probability that any comparison will be incorrectly found significant (Hochberg and Tamhane, 1987)

$$\Sigma = \frac{\sigma}{\sqrt{n}} \text{ for very big population sizes (1)}$$

However, this formula assumes that the population N is very great and that the n/N is very small. In the CPI, if one accepts that the population size is just 13, that is the maximum number of sources that could have been used to evaluate a country, then the assumptions for the formula of the standard error above do not hold. Instead, the correct formula to be used can be found in the seminal work of Isserlis (1918), where the standard error of the mean is:

$$\Sigma = \sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}} \text{ for small population sizes (2)}$$

Hence, we recommend that the standard errors for the CPI scores are calculated using the formula for small population sizes. The standard errors calculated with formula (2) are 9% less than the standard errors obtained with formula (1) for countries that were evaluated by three sources, up to 50% less for countries that were evaluated by ten sources.

After these considerations, we assess whether there is a pattern between the standard errors associated with the CPI scores and the number of sources that were used to evaluate a country. We will calculate the standard errors using the formula (2) above for small population sizes. Again, we apply a multiple comparison test after Bonferroni correction for the group means of the standard errors. The results suggest that overall there is a negative association between the standard errors and the number of sources, implying that standard errors calculated over a small number of sources are greater (on average) that standard errors calculated over many sources. Yet, the pattern is not linear. To be more specific, standard errors calculated over three sources are not different (on average) from those calculated over four or five sources, but are significantly greater than those calculated over six or more sources. Similarly, standard errors calculated over four sources are significantly greater than those calculated over seven or more sources, but no different than those calculated over five or six sources. Interestingly, standard errors calculated over ten sources[5] are not significantly different (on average) that those calculated over six or more sources, but are significantly lower than those calculated over three to five sources. This result suggests that the criterion for a country's inclusion to the CPI could have been more conservative, from three sources (currently) to six

[5] Ten is the maximum number of sources based on which a country is evaluated in the CPI 2012.

sources, in order to avoid potential criticism that countries evaluated on three to five sources have more uncertain CPI scores. Yet, introducing such a conservative criterion would imply leaving 58 countries outside the CPI. Even in that case, a counterargument would be that given that standard errors calculated over six sources are not significantly different from those calculated over four or more sources, the criterion could be relaxed to the inclusion in the CPI of countries that are evaluated by at least four sources. And even then, given that the standard errors calculated over four sources are not significantly different from those calculated over three sources, the current criterion for a country's inclusion in the CPI seems reasonable. Altogether, these results suggest that there seems to be no bias in the CPI scores with respect to the number of sources used, whilst countries with few available sources tend to have slightly larger standard errors (on average) compared to countries that are evaluated using more sources. Nevertheless, the statistical analysis does not provide a clear suggestion as to whether the criterion for a country's inclusion in the CPI if evaluated by at least three sources should be modified.

Table 4. Multiple comparison: means of CPI standard errors grouped by the number of sources

Number of sources	3	4	5	6	7	8	9
4	no						
5	no	no					
6	yes	no	no				
7	yes	yes	yes	no			
8	yes	yes	yes	no	no		
9	yes	yes	yes	no	no	no	
10	yes	no	yes	no	no	no	no

Source: Saisana and Saltelli, 2012, European Commission Joint Research Centre

Notes: A multiple comparison test after Bonferroni correction was applied. For the comparison 3-4, "no" implies that the group mean of standard errors for countries evaluated on three sources is not significantly different (at 5% level) from the group mean of standard errors for countries evaluated on four sources.

4. Interpreting the CPI rankings: effect size

The 2012 CPI scores are reported at two digits and are accompanied by a standard error of estimate and the 90% confidence interval. Afghanistan, Korea (North) and Somalia score 8 points, which is the highest perceived level of corruption, whilst Denmark, Finland and New Zealand score 90 points, which is the lowest level of perceived corruption among the 176 countries analysed. Yet, is the level of perceived corruption different in countries with 1 or 2 points difference in their CPI scores? To interpret the difference between two countries scores, we employ the *effect size*. The effect size is a simple way to quantify the difference between two countries without confounding the interpretation with the sample size, as is the case in the statistical significance (see Section 2). There is a wide array of formulas used to measure effect size. We used *Cohen's d* formula (Cohen, 1988; Hartung et al., 2008; Hedges, 1981) for two countries:

$$\text{effect size} = \frac{(M_1 - M_2)}{\sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}}}$$

M_1 and M_2 refer to the CPI country scores, N_1 and N_2 are the number of sources available for each country, SD_1 and SD_2 are the standard deviations across the sources that were used to evaluate each country. Country 1 is the highest ranked country in the comparison. The denominator in the equation above is a so-called 'pooled' estimate of the standard deviation for both countries. Essentially this estimate is an average of both standard deviations[6]. Cohen (1988) hesitantly defined effect sizes as "*small*, threshold =0.2", "*medium*, threshold = 0.5", and "*large*, threshold = 0.8"[7]. These effect sizes correspond respectively to a non-overlap of 14.7%, 33.0% and 47.4% in the two distributions. Effect sizes smaller than 0.2 suggest that there may be no difference in the average country scores given the large overlap in the two distributions.

[6] Note that this 'pooled' estimate does not equal the standard deviation of the 'pooled' data set, i.e. the data set including the values of both countries. If both countries have a low standard deviation but show a big difference in average score, the latter estimate will be much bigger than the true pooled estimate of the standard deviation

[7] Cohen (1988) stated that "there is a certain risk inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science" (p.25)

Table 5 gives the effect size of the differences in the CPI scores between any two countries in the top 20 (those with the least perceived level corruption). The CPI scores for the first three countries −Denmark, Finland and New Zealand− do not show a significant difference between them (small effect sizes of less than 0.3). Hence, the CPI rank 1 has been correctly assigned to these three countries. Results confirm that these three countries are better off than all the remaining countries. Sweden and Singapore have an effect size of 0.6, which is equivalent to a non-overlap in their distributions of 38%. Depending on which threshold value is chosen for the effect size, they could either be placed on equal footing or place Sweden higher than Singapore. Further down in the CPI ranking, Belgium (rank 16), Japan and United Kingdom (both at rank 17) could actually be placed on equal footing. Similarly, the United States (rank 19) could actually be considered as having the same level of perceived corruption as Chile and Uruguay (rank 20).

The largest effect size of 1.8 in the top five countries arises when New Zealand and Singapore are compared. This indicates that the average score for New Zealand is significantly higher than the average score for Singapore, but that there is an overlap of 23% in the two distributions that should not be ignored. For comparison, the group of top performers in the Global Insight Country Risk Ratings includes nine countries − Denmark, Finland, New Zealand, Sweden, Singapore, Australia, Norway, Canada, and Netherlands− that are all in the top 10 of the CPI classification. Interestingly, Qatar and Switzerland have the same level of perceived corruption according to the Global Insight Country Risk Ratings, but are significantly different in their CPI scores. In fact their CPI scores have an effect size over 3, implying that there is no overlap in the two distributions. These results show that the CPI −by taking into account a plurality of sources− suggests that the average level of perceived corruption is different in those countries, unlike what the Country Risk Ratings suggest.

Overall, the CPI ranking accurately reflects when country differences are significant and when not. Yet, it is important that even significant differences are carefully interpreted given that there might be a substantial overlap in the resulting distributions for the countries.

Table 5. Top twenty CPI scores: Effect sizes of pairwise country comparisons

CPI Rank	N	SD	CPI Score	Country	Denmark	Finland	New Zealand	Sweden	Singapore	Switzerland	Australia	Norway	Canada	Netherlands	Iceland	Luxembourg	Germany	Hong Kong	Barbados	Belgium	Japan	United Kingdom	United States	Chile
1	7	1.9	90	Denmark																				
1	7	2.9	90	Finland	0.0																			
1	7	2.1	90	New Zealand	-0.3	-0.2	0.0																	
4	7	1.9	88	Sweden	1.1	0.9	1.3	0.0																
5	9	2.1	86	Singapore	1.6	1.3	1.8	0.6	0.0															
6	6	2.6	86	Switzerland	1.8	1.5	2.0	0.8	0.3	0.0														
7	8	1.2	85	Australia	3.2	2.4	3.4	1.9	1.1	0.5	0.0													
7	7	1.6	85	Norway	2.9	2.2	3.1	1.7	1.0	0.5	0.0	0.0												
9	7	2.1	84	Canada	3.0	2.4	3.2	2.0	1.4	0.9	0.6	0.5	0.0											
9	7	2.0	84	Netherlands	2.7	2.2	2.9	1.7	1.0	0.6	0.2	0.2	-0.3	0.0										
11	6	4.1	82	Iceland	2.5	2.2	2.6	1.8	1.5	1.1	0.9	0.9	0.5	0.7	0.0									
12	6	2.9	80	Luxembourg	4.0	3.3	4.1	3.1	2.6	2.0	2.2	2.0	1.4	1.7	0.5	0.0								
13	8	2.3	79	Germany	5.0	4.2	5.1	4.1	3.5	2.8	3.1	2.9	2.1	2.5	1.0	0.5	0.0							
14	8	2.0	77	Hong Kong	6.3	5.1	6.3	5.3	4.5	3.7	4.5	4.0	3.0	3.5	1.5	1.2	0.7	0.0						
15	3	6.5	76	Barbados	3.8	3.4	3.9	3.3	3.1	2.4	2.8	2.5	2.1	2.3	1.3	1.1	0.9	0.5	0.0					
16	7	2.4	75	Belgium	6.9	5.7	6.9	6.0	5.3	4.4	5.4	4.9	3.9	4.3	2.2	2.1	1.8	1.3	0.3	0.0				
17	9	2.4	74	Japan	7.2	6.0	7.2	6.3	5.6	4.7	5.6	5.2	4.2	4.6	2.5	2.4	2.1	1.6	0.5	0.2	0.0			
17	8	1.3	74	Un. Kingdom	9.5	7.1	9.3	8.3	7.0	5.8	8.4	7.3	5.4	6.0	2.8	2.9	2.5	1.9	0.5	0.2	-0.1	0.0		
19	9	4.0	73	United States	5.2	4.8	5.3	4.5	4.3	3.7	3.9	3.7	3.2	3.5	2.3	2.0	1.8	1.4	0.6	0.5	0.4	0.4	0.0	
20	9	2.1	72	Chile	8.7	7.2	8.7	7.7	6.9	5.9	7.3	6.7	5.5	5.9	3.3	3.4	3.1	2.6	1.1	1.1	0.9	1.2	0.2	0.0
20	6	1.5	72	Uruguay	10.0	7.5	9.7	8.9	7.6	6.4	9.5	8.1	6.1	6.7	3.2	3.5	3.4	3.0	1.0	1.2	1.0	1.5	0.2	0.0

Source: Saisana and Saltelli, 2012, European Commission Joint Research Centre

5. Impact of modelling assumptions on the CPI

Robustness of the CPI with respect to its imputation and normalisation scheme

As described in Section 2, the CPI 2012 is calculated as the simple average of standardized scores across the available sources for each country. A related concern is whether the CPI ranking is sufficiently robust to the choice of the 'global' mean and standard deviation that were estimated using the "impute" command in STATA. To test this, we apply an Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977; Little and Rubin, 1992) in the statistical software MATLAB to estimate the 'global' mean and standard deviation for each source.[8] The simulated country scores were then calculated using a simple average of the standardized scores (only those that were available per country).

The results show that the CPI ranking and the simulated ranking are very similar: the Spearman rank correlation is 0.9987. There is absolutely no difference between the CPI ranking and the simulated ranking for 52 countries, whilst there is less than five-rank difference for 165 countries (94% of the cases). These results demonstrate that the CPI 2012 ranking is robust to the estimation of the 'global' parameters (mean and standard deviation) which are subsequently used to render the scores from the thirteen sources comparable.

Evaluating each source's contribution to the final CPI score

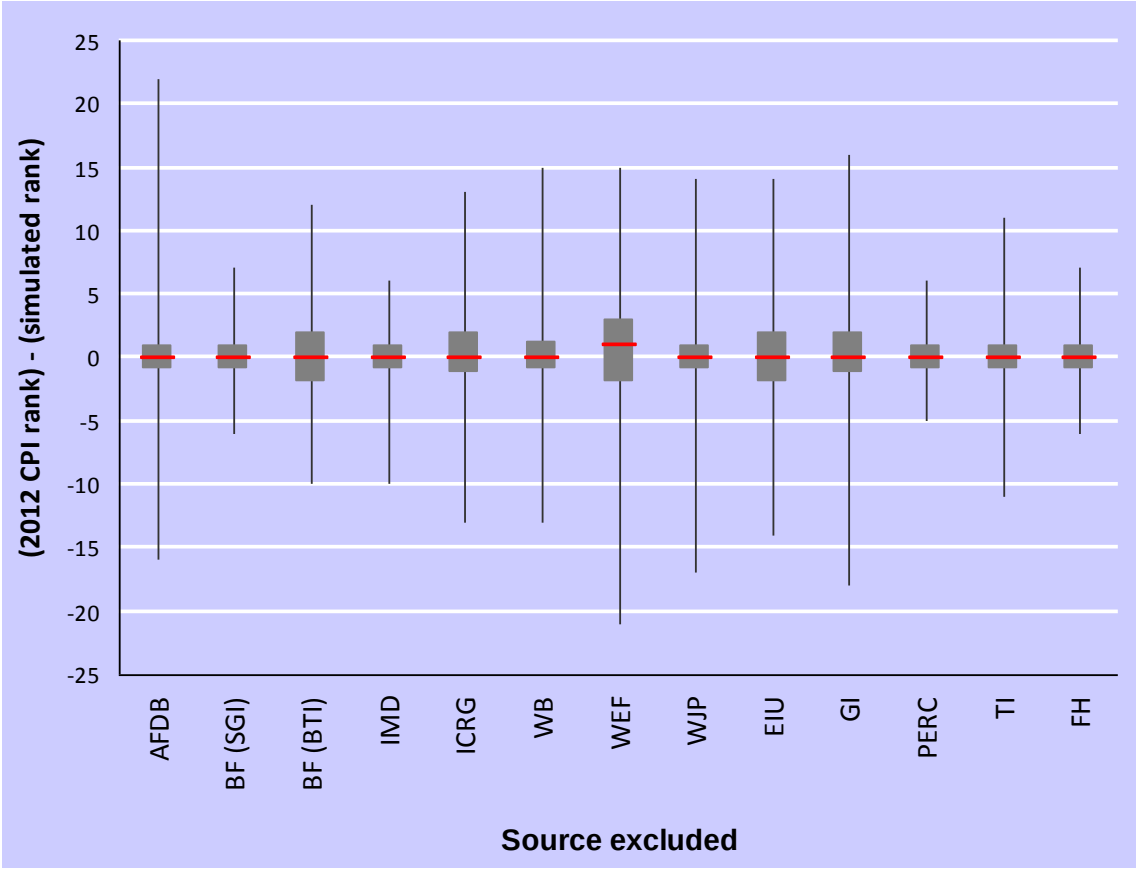
A further concern relates to whether the CPI is well balanced across the thirteen sources of perceived corruption. In other words, are all sources equally important in determining the CPI ranking? If the country coverage for each source was at least 50, we would have calculated the importance of each source using a non-linear measure, the kernel estimate of the Pearson correlation ratio[9]. Instead, given that some sources have very limited country coverage, we tested

[8] The EM algorithm is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating the following steps: (a) the expectation E-step: given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates. (b) The maximization M-step: given complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step.

[9]Paruolo et al., 2013, discuss four properties of the Pearson correlation ratio (else termed first order sensitivity measure), which render the correlation ratio a suitable measure of the indicators' importance: (a) it offers a precise definition of importance, that is 'the expected reduction in variance of an index that would be obtained if a

the impact of each source on the CPI ranking by excluding a source at a time and comparing the shift in rank with respect to a CPI rank. We did so only for countries that were evaluated by at least four sources, so that by excluding a source a country is evaluated by at least three source (criterion for inclusion in the CPI).

Figure 2. CPI framework: Impact of excluding a source



Source: Saisana and Saltelli, 2012, European Commission Joint Research Centre

The main results are provided in Figure 2. The red line is the median across all countries and the boxes include 50 percent of the cases. The whole distribution of the rank differences is displayed by the vertical blue lines. A median close to zero with a small box and a short vertical line indicates a source whose exclusion does not affect significantly the final rank. For all sources, the median is close to zero and the box is within ± 2 positions, which suggests that eliminating any of the sources would practically leave unaffected half of the countries. For some of the remaining countries, the most influential sources in determining their CPI rank are the African

variable could be fixed'; (b) it can be used regardless of the degree of correlation between variables; (c) it is model-free, in that it can be applied also in non-linear aggregations; (d) it is not invasive, in that no changes are made to the index or to the correlation structure of the indicators.

Development Bank Governance Ratings, the World Economic Forum Executive Opinion Survey, the World Justice Project Rule of Law Index, and the Global Insight Country Risk Ratings. Yet, the influence is moderate for the majority of the countries. In fact, the maximum shift with respect to the CPI rank when excluding a source is up to 4 positions for 75 percent of the countries. This suggests that no source dominates the overall index and that all sources contribute to determining the CPI ranking in a balanced way.

6. Conclusions

The JRC analysis suggests that the new methodology for the Corruption Perceptions Index (CPI), besides being appealing for reasons of transparency and replicability, it is also conceptually and statistically coherent and with a balanced structure (i.e., the CPI is not dominated by any of the individual sources). Despite the high associations between the sources, the information offered by the CPI is shown to be non redundant. There seems to be no bias in the CPI scores with respect to the number of sources used, whilst countries with few available sources tend to have slightly larger standard errors (on average) compared to countries that are evaluated using more sources. Results also provided statistical justification for the use of simple average across the sources. Country ranks are in most cases fairly robust to the key assumption on the estimation of global parameters (mean and standard deviation) for each source.

Altogether, the statistical analyses described in this report underline the contribution of the CPI to the measurement of perceived corruption in the public sector at national level worldwide:

- the CPI covers more countries than any of the individual sources alone,
- the CPI may be more reliable than each source taken separately,
- the CPI can efficiently differentiate the level of corruption between countries, unlike some sources where a large number of countries is assessed at the same level of corruption,
- the CPI reconciles different view points on the issue of corruption, noteworthy since no country is classified as better off than another country on all common sources.

The main recommendation for the CPI team is to adjust the formula for the standard errors for the small population size and for policy makers to consider the statistical significance (by means

of effect size for example) when comparing the CPI scores. The results make clear that even when differences in the CPI country scores are statistically significant they should be carefully interpreted.

References

Cohen, J. (1969) *Statistical power analysis for the behavioral sciences*. NY: Academic Press.

Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society. B 39 (1): 1–38.

Hartung, J., Knapp G., Sinha B.K. (2008) *Statistical Meta-Analysis with Application*. Hoboken, New Jersey: Wiley.

Hedges, L.V. (1981) Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6 (2): 107–128.

Hochberg Y., Tamhane A.C. (1987) Multiple Comparison Procedures, Wiley.

Isserlis, L. (1918) On the value of a mean as calculated from a sample, *Journal of the Royal Statistical Society*, 81:75-81.

Little, R. J. A., Rubin, D. B. (2002) S*tatistical Analysis with Missing Data*, 2nd edition. Hoboken, NJ: John Wiley & Sons.

OECD/EC JRC (Organisation for Economic Co-operation and Development / European Commission Joint Research Centre). 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD.

Paruolo, P., Saisana, M., Saltelli, A. (2013) Ratings and rankings: Voodoo or Science? *Journal of the Royal Statistical Society A,* 176 (2):1-26.

Saisana, M., B. D'Hombres, and A. Saltelli (2011) Rickety Numbers: Volatility of University Rankings and Policy Implications. *Research Policy* 40: 165–77.

Saisana, M., A. Saltelli, and S. Tarantola. (2005) 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators. *Journal of the Royal Statistical Society* A 168 (2): 307–23.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. *Global Sensitivity Analysis: The Primer*. Chichester, England: John Wiley & Sons.

European Commission
EUR 25623 EN– Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **Corruption Perceptions Index 2012 - Statistical Assessment**

Authors: Michaela Saisana, Andrea Saltelli

Luxembourg: Publications Office of the European Union

2012 – 26 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-27732-0

doi:10.2788/69609

Abstract
The Corruption Perceptions Index (CPI) by Transparency International measures perceptions of corruption in the public sector in different countries around the world. Upon invitation of the CPI team, the JRC assessed the new methodology in the CPI 2012 and analyzed the consequences that come with this change. The statistical assessment of the Index was done along three main avenues: an evaluation of conceptual/statistical coherence of the index structure, an interpretation of the rankings based on significance tests, and an evaluation of the impact of key modelling assumptions (imputation and normalisation) on countries' scores and ranks. The CPI 2012 passes all the statistical filters of the quality control. The main recommendation for the CPI team is to adjust the formula for the standard errors for the small population size (errors that are currently overestimated) and for policy makers to consider the statistical significance (by means of effect size for example) when comparing the CPI scores. The results make clear that even when differences in the CPI country scores are statistically significant they should be carefully interpreted.

LB-NA-25623-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.