European Commission

JRC TECHNICAL REPORTS

Corruption Perceptions Index 2017 Statistical Assessment

Álvarez-Díaz, Marcos
Saisana, Michaela
Montalto, Valentina
Tacao Moura, Carlos

2018

Joint Research Centre

EUR 29405 EN

Contact information
Competence Centre on Composite Indicators and Scoreboards
jrc-coin@ec.europa.eu

JRC Science Hub
https://ec.europa.eu/jrc/en/coin
https://composite-indicators.jrc.ec.europa.eu/

JRC113251

EUR 29405 EN

PDF ISBN 978-92-79-96745-0 ISSN 1831-9424 doi: 10.2760/974516

Luxembourg: Publications Office of the European Union, 20018

How to cite this report: Alvarez Diaz, M., Saisana, M., Montalto, V. and Tacao Moura, C., Corruption Perceptions Index 2017 Statistical Assessment , EUR 29405 EN, Publications Office of the European Union, Luxembourg, 2018, ISBN 978-92-79-96745-0 (online), doi:10.2760/974516 (online), JRC113251.

Contents

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 others.

The European Commission's Competence Centre on Composite Indicators and Scoreboards at the Joint Research Centre (JRC) in Ispra, Italy, was invited by Transparency International to assess the statistical properties of the CPI 2017. This audit represents the second analysis of the CPI since a first one was conducted in 2012.

As in the previous audit, 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 2017 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 statistical coherence of the CPI 2017 is based on an analysis of the covariance structure across different 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, although countries evaluated on three and four sources present more uncertain scores. A recommendation is made on the calculation of the standard errors, which are overestimated by the current formula used by the developers of the CPI.

The modelling assumptions (normalisation coupled with estimation of missing data) are found to have a moderate impact on the CPI ranking (no impact for 34 countries, less than or equal to five-rank shift for 84 % of the countries). When one of the sources is

excluded, the analysis also shows that the ranking shift with respect to the CPI rank is more than three positions for a range between 5 % and 34 % of the countries, depending on the source deleted. This fact suggests that all sources contribute, to a greater or lesser extent, to determining the CPI ranking.

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:

- After Global Insight Country Risk Ratings 2016, 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 (e.g. all countries ranked in the Global Insight Country Risk Ratings 2016 are tied with some others);

- The CPI reconciles different point of views 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 policymakers 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 22 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 others, as done in the CPI, is both advantageous but also potentially worrisome. The main advantage and added value of the CPI lies 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's Competence Centre on Composite Indicators and Scoreboards at the Joint Research Centre (JRC) in Ispra, Italy, was invited by Transparency International to assess the statistical properties of the CPI 2017. 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 analysed the revised methodology of the CPI 2017 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 2017 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 is structured as follows.

Section 2 presents the 13 sources that were used in the CPI 2017, as well as the methodology used to construct the index.

([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/.

Section 3 analyses the statistical coherence of the CPI 2017 based on an analysis of the covariance structure across the 13 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 used by the developers.

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 policymakers 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 modelling assumptions (normalisation 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 34 countries, whilst there is a less than or equal to five-rank difference for 84 % of the countries.

The analysis also shows that the shift with respect to the CPI rank when excluding one of the sources is more than three positions from 5 % to 34 % of the countries; the percentage depends on the source excluded. Moreover, the CPI 2017 has a very high statistical reliability (it has a Cronbach's alpha value of 0.92), and it is not strongly affected when one source is deleted at a time. All these facts suggest that all sources contribute, to a greater of lesser extent, 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 has been an evolving project since 1995. Every year, such measurement builds upon previous editions while refined with newly available data. The CPI 2017 is calculated for 180 countries around the world, and it is based on 13 sources that collect the assessment of experts and business executives on some specific corrupt behaviour in the public sector (i.e. bribery, diversion of public funds, use of public office for private gain, nepotism in the civil service and state capture). The sources of information used to build the CPI are listed in Table 1. The sources differ in the number of countries covered, ranging from 15 countries covered in the Political and Economic Risk Consultancy Asian Intelligence to 180 countries included in the Global Insight Country Risk Ratings. The source Varieties of Democracy was included for the first time in 2016, to the detriment of the source Transparency International Bribe Payers Survey. More detailed information on the sources and the rationale for inclusion of each source is offered in the main report of the CPI 2017.

The most recently released country scores from those 13 sources were used in the development of the CPI 2017. Countries were included if they were evaluated by at least three sources; this was the case for 13 countries (e.g. Barbados, Bahamas, Grenada). The maximum number of sources based on which a country was evaluated was 10; this was the case for nine countries (i.e. Bulgaria, Croatia, Czechia, Estonia, Hungary, Poland, Romania, Slovenia and South Korea). Most countries were evaluated using seven (35 countries) and eight sources (38 countries).

For simplicity in communication and to allow comparisons over time, the CPI 2017 is calculated using a simple average of standardised scores. More specifically, all 13 sources are standardised 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.

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

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 four sources: Economist Intelligence Unit Country Risk Ratings, Freedom House Nations in Transit, Political and Economic Risk Consultancy Asian Intelligence, and Varieties of Democracy Project's Political Corruption Index.

After the standardisation, any values beyond the 0-100 scale are capped. For the normalised scores to be comparable between the 13 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 have been if all 180 countries had been evaluated by each source? As in previous editions, the CPI 2017 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 180 countries are then calculated and used as the parameters to standardise the sources during the normalisation. 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 normalised scores across the available sources only. The CPI scores are in the range 0 to 100 (0 being the lowest level of perceived corruption).

Table 1. *2017 CPI Sources of information and number of countries in common with the CPI*

Source	**Number of countries**
1. African Development Bank Governance Ratings (AFDB) 2016	38
2. Bertelsmann Stiftung Governance Indicators (BF-SGI) 2017	41
3. Bertelsmann Stiftung Transformation Index 2017-2018 (BF-BTI)	129
4. Economist Intelligence Unit Country Risk Service (EIU) 2017	131
5. Freedom House Nations in Transit (FH) 2017	29
6. Global Insight Country Risk Ratings (GI) 2016	180
7. IMD World Competitiveness Center World Competitiveness Yearbook Executive Opinion Survey (IMD) 2017	63
8. Political and Economic Risk Consultancy Asian Intelligence (PERC) 2017	15
9. The PRS Group International Country Risk Guide (ICRG) 2017	140
10. World Bank — Country Performance and Institutional Assessment (WB) 2017	67
11. World Economic Forum Executive Opinion Survey (WEF) 2017	133
12. World Justice Project Rule of Law Index Expert Survey (WJP) 2017-2018	110
13. Varieties of Democracy Project's Political Corruption Index (V-Dem) 2017	169

Source: Corruption Perceptions Index 2017.

3 Conceptual and statistical coherence in the CPI

Each of the 13 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 13 sources taken independently.

Assessing potential redundancy of information in the CPI

The country rankings from the 13 different sources tend to correlate well with each other. There is also a high correlation between the CPI ranking and each of the sources, ranging from 0.87 to 0.95 (see Table 3). These high correlations were expected, given that all sources attempt to measure the same phenomenon, which 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 13 sources cover different countries — from 15 countries for the Political and Economic Risk Consultancy Asian Intelligence to 180 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, 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 to have the same perceived level of corruption (e.g. while the Global Insight Country Risk Ratings only has seven different scores for 180 countries, the CPI presents 66 different scores for the same number of countries). One more feature of the CPI is that it reconciles different viewpoints on the issue of corruption. If the countries' classifications in the 13 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 WEF, GI, BF-BTI, PRS, VDEM and EIU (78 countries are common to all sources) ([2]). The first latent dimension accounts for 80 % 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 are 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	4.8	79.5	WEF	0.87
2	0.7	91.6	GI	0.96
3	0.2	95.4	BF (BTI)	0.66
4	0.2	97.9	PRS	0.96
5	0.1	99.2	VDEM	0.91
6	0.0	100.0	EIU	0.96

Source: Own elaboration.

([2]) PCA could not be applied to the entire set of 13 sources because no single country is covered by all sources.
([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	WB	WEF	GI	BF-BTI	AFDB	IMD	BF-SGI	WJP	ICRG	V-DEM	EIU	FH	PERC
CPI	-	0.87	0.68	0.90	0.75	0.67	0.79	0.83	0.83	0.87	0.77	0.95	0.81	0.85
WB	0.87 (n=67)	—	0.36	0.74	0.74	0.79	—	—	0.73	0.74	0.60	0.61	0.75	—
WEF	0.84 (n=133)	0.39 (n=39)	-	0.68	0.36	0.38	0.83	0.65	0.65	0.67	0.52	0.73	0.32	0.75
GI	0.92 (n=180)	0.66 (n=67)	0.75 (n=133)	-	0.73	0.42	0.72	0.84	0.84	0.82	0.75	0.92	0.79	0.95
BF-BTI	0.86 (n=129)	0.72 (n=52)	0.46 (n=100)	0.74 (n=129)	-	0.70	0.30	0.85	0.55	0.70	0.57	0.79	0.91	0.71
AFDB	0.81 (n=38)	0.82 (n=37)	0.45 (n=25)	0.44 (n=38)	0.76 (n=33)	-	-	-	0.59	0.62	0.34	0.57	-	-
IMD	0.91 (n=63)	- (n=1)	0.95 (n=62)	0.79 (n=63)	0.37 (n=37)	- (n=0)	-	0.72	0.73	0.77	0.65	0.83	0.44	0.79
BF-SGI	0.88 (n=41)	- (n=0)	0.74 (n=40)	0.80 (n=41)	0.84 (n=15)	- (n=0)	0.80 (n=40)	—	0.83	0.76	0.79	0.75	0.95	—
WJP	0.94 (n=110)	0.74 (n=33)	0.80 (n=93)	0.88 (n=110)	0.68 (n=81)	0.72 (n=16)	0.88 (n=51)	0.88 (n=31)	-	0.71	0.73	0.82	0.56	0.89
ICRG	0.94 (n=140)	0.68 (n=37)	0.80 (n=115)	0.84 (n=140)	0.76 (n=105)	0.66 (n=26)	0.89 (n=63)	0.80 (n=41)	0.80 (n=97)	-	0.75	0.91	0.91	0.66
V-DEM	0.91 (n=169)	0.65 (n=63)	0.68 (n=128)	0.81 (n=169)	0.70 (n=127)	0.47 (n=38)	0.80 (n=60)	0.82 (n=39)	0.88 (n=)103	0.86 (n=130)	-	0.82	0.77	0.78
EIU	0.93 (n=131)	0.43 (n=28)	0.79 (n=111)	0.87 (n=131)	0.76 (n=100)	0.52 (n=16)	0.86 (n=63)	0.72 (n=41)	0.83 (n=95)	0.88 (n=123)	0.83 (n=126)	-	0.76	0.89
FH	0.92 (n=29)	- (n=5)	0.43 (n=23)	0.80 (n=29)	0.96 (n=29)	- (n=0)	0.53 (n=14)	0.93 (n=11)	0.69 (n=20)	0.92 (n=20)	0.89 (n=29)	0.74 (n=23)	-	-
PERC	0.95 (n=15)	- (n=2)	0.90 (n=15)	0.95 (n=15)	0.82 (n=11)	- (n=0)	0.92 (n=13)	- (n=4)	0.96 (n=14)	0.78 (n=14)	0.91 (n=14)	0.92 (n=15)	- (n=0)	-

Source: Own elaboration.

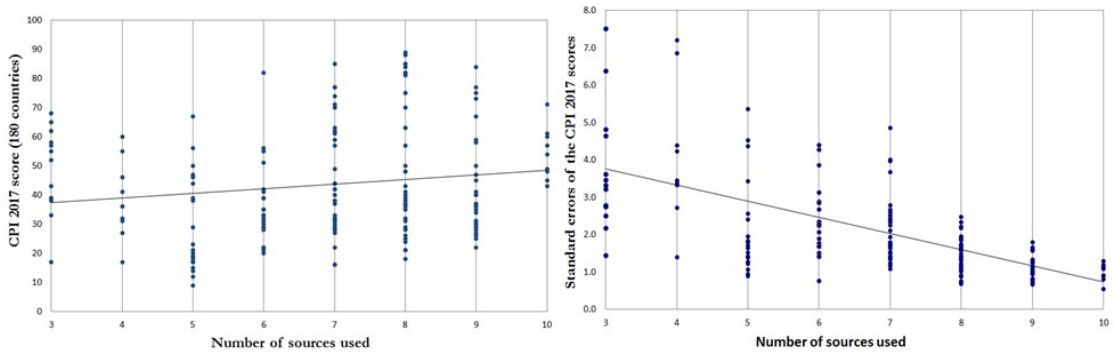
NB: Low diagonal: Spearman rank correlation coefficients (significant at 5 % level). Number of countries that are common to each pair of sources is given in the parenthesis. Upper diagonal: Gamma statistic (significant at the 5 % level), which is to be preferred over the Spearman rank correlation for sources with tied values. 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 13 countries, up to 10 sources that were used to evaluate nine countries, see Figure 1(a)). 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 three, four, 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*

(a) *Impact on the CPI scores* (b) *Impact on the standard errors*



Source: Own elaboration.

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:

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

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

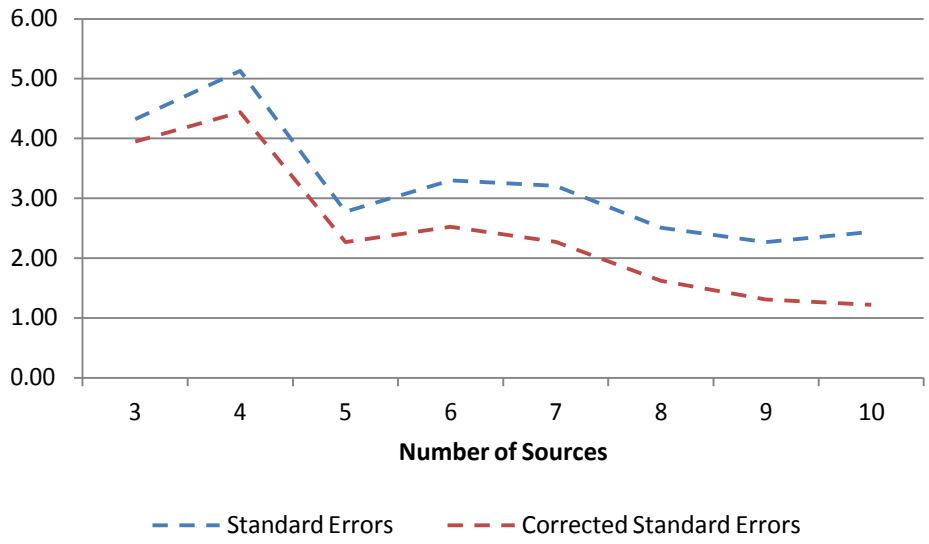
([4]) When performing a simple t-test of one group mean against another, one needs to specify a significance level that determines the cutoff value of the t-statistic. For example, one can specify the value alpha = 0.05 to ensure that when there is no real difference, one will incorrectly find a significant 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).

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. Figure 2 compares the standard errors calculated according to the formula (1), and the correction calculated with formula (2). The corrected standard errors show lowest values regardless of the number of sources. In fact, the corrected standard errors are 9 % less than the standard errors obtained with the formula (1) for countries that were evaluated by three sources, up to 50 % less for countries that were evaluated by 10 sources.

Figure 2. *Comparison between the standard errors and the corrected standard errors grouped by sources.*



Source: Own elaboration.

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. Figures 1(b) and 2 also 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) than standard errors calculated over many sources. Additionally, Table 4 presents the results of the multiple comparison tests after Bonferroni correction for the group means of the

standard errors calculated using the formula (2) above for small population sizes. To be more specific, standard errors calculated over three sources are not different (on average) from those calculated over four sources; but the standard errors associated to both sources are significantly greater than those calculated over five or more sources. This result suggests the criterion for a country's inclusion to the CPI could have been more conservative, from three sources (currently) to five, in order to avoid potential criticism that countries evaluated on three and four sources have more uncertain CPI scores. Yet, introducing such a conservative criterion would imply leaving 22 countries outside the CPI.

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	YES	YES					
6	YES	YES	NO				
7	YES	YES	NO	NO			
8	YES	YES	NO	YES	NO		
9	YES	YES	YES	YES	YES	NO	
10	YES	YES	NO	YES	NO	NO	NO

Source: Own elaboration.
NB: A multiple comparison test after Bonferroni correction was applied. Reading: 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 CPI 2017 scores are reported at two digits and are accompanied by a standard error of estimate and the 90 % confidence interval. The highest perceived levels of corruption are registered for Somalia (9 points), Sudan (12 points), Syria (14 points) and Afghanistan (15 points). Conversely, the lowest levels of perceived corruption among the 180 countries analysed are for New Zealand (89 points), Denmark (88) and Finland, Norway and Switzerland (the latter countries with 85 points each). Yet, is the level of perceived corruption different in countries with one or two 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. 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:

$$\text{effect size} = \frac{(M_1 - M_2)}{\sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2}{N_1 + N_2 - 2}}} \quad (3)$$

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 ([5]). Cohen (1988) hesitantly defined effect sizes as '*small*, threshold = 0.2', '*medium*, threshold = 0.5', and '*large*, threshold = 0.8' ([6]). 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.

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 two countries — New Zealand and Denmark — do not show a significant

([5]) 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.

([6]) 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 field of inquiry as behavioral science' (p.25).

difference between them. They show a medium effect size, which is equivalent to a non-overlap in their distribution of less than 33 %. Results confirm that these two countries are better off than all the remaining countries. Finland, Norway and Switzerland do not show differences in the average country given that their effect sizes are equal to zero (in fact, they perform the same score, 85 points). These three countries present an effect size less than or equal to 0.5 in comparison with Singapore and Sweden. Sweden and Singapore are on equal footing, and better off than the rest of the countries. Further down in the CPI ranking, Ireland (rank 19) and Japan (rank 20) could actually be considered to have the same level of perceived corruption.

The largest effect size of 1.9 in the top five countries arises when New Zealand and Switzerland are compared. This indicates that the average score for New Zealand is significantly higher than the average score for Switzerland, but that there is an important overlap in the two distributions that should not be ignored. For comparison, the group of top performers in the Global Insight Country Risk Ratings includes 15 countries — New Zealand, Denmark, Finland, Norway, Switzerland, Singapore, Sweden, Canada, Luxembourg, Netherlands, United Kingdom, Germany, Hong Kong, Iceland and Belgium. Except Belgium, all of them are in the top 15 of the CPI classification. Interestingly, New Zealand and Hong Kong have the same level of perceived corruption according to the Global Insight Country Risk Ratings (83 points), but are significantly different in their CPI scores. In fact, their CPI scores have an effect size of 5.0, implying that there is no overlap in the two distributions. Therefore, 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 Global Insight Country Risk Ratings suggests.

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 20 CPI scores: Effect sizes of pairwise country comparisons.

CPI Rank	*N*	*SD*	*CPI Score*	*Country*	*New Zealand*	*Denmark*	*Finland*	*Norway*	*Switzerland*	*Singapore*	*Sweden*	*Canada*	*Luxembourg*	*Netherlands*	*UK*	*Germany*	*Australia*	*Hong Kong*	*Iceland*	*Austria*	*Belgium*	*USA*	*Ireland*	*Japan*
1	8	2.4	89	New Zealand	0.0																			
2	8	2.8	88	Denmark	0.4	0.0																		
3	8	2.8	85	Finland	1.5	1.1	0.0																	
3	8	1.9	85	Norway	1.8	1.3	0.0	0.0																
3	7	1.7	85	Switzerland	1.9	1.3	0.0	0.0	0.0															
6	9	2.3	84	Singapore	2.1	1.6	0.4	0.5	0.5	0.0														
6	8	2.3	84	Sweden	2.1	1.6	0.4	0.5	0.5	0.0	0.0													
8	8	1.5	82	Canada	3.4	2.7	1.3	1.8	1.9	1.0	1.0	0.0												
8	6	2.1	82	Luxembourg	3.0	2.4	1.2	1.5	1.6	0.9	0.9	0.0	0.0											
8	8	2.3	82	Netherlands	2.9	2.3	1.2	1.4	1.5	0.9	0.9	0.0	0.0	0.0										
8	8	1.7	82	UK	3.3	2.6	1.3	1.7	1.8	1.0	1.0	0.0	0.0	0.0	0.0									
12	8	1.9	81	Germany	3.7	2.9	1.7	2.1	2.2	1.4	1.4	0.6	0.5	0.5	0.5	0.0								
13	9	1.4	77	Australia	6.1	5.0	3.7	4.9	5.2	3.7	3.8	3.4	2.9	2.6	3.2	2.4	0.0							
13	7	2.4	77	Hong Kong	5.0	4.2	3.0	3.8	3.9	3.0	3.0	2.5	2.2	2.1	2.4	1.9	0.0	0.0						
13	7	4.3	77	Iceland	3.5	3.1	2.2	2.5	2.4	2.1	2.1	1.6	1.4	1.5	1.6	1.2	0.0	0.0	0.0					
16	8	1.2	75	Austria	7.3	6.1	4.6	6.4	7.0	4.9	5.0	5.2	4.3	3.8	4.7	3.8	1.5	1.1	0.7	0.0				
16	8	1.6	75	Belgium	6.8	5.7	4.4	5.8	6.2	4.6	4.6	4.5	3.8	3.5	4.2	3.4	1.3	1.0	0.6	0.0	0.0			
16	9	3.2	75	USA	4.8	4.3	3.3	3.7	3.7	3.2	3.2	2.7	2.4	2.5	2.6	2.2	0.8	0.7	0.5	0.0	0.0	0.0		
19	7	3.6	74	Ireland	4.9	4.4	3.4	3.9	3.9	3.4	3.4	3.0	2.6	2.7	2.9	2.5	1.2	1.0	0.8	0.4	0.4	0.3	0.0	
20	9	2.7	73	Japan	6.3	5.5	4.4	5.2	5.2	4.5	4.4	4.1	3.6	3.6	4.0	3.4	1.9	1.6	1.2	1.0	0.9	0.7	0.3	0.0

Source: Own elaboration.

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 2017 is calculated as the simple average of standardised 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 ([7]). The simulated country scores were then calculated using a simple average of the standardised 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 0.995. Table 6 reports the differences in the ranking according to the method of imputation used in the normalisation process. There is absolutely no difference between the CPI ranking and the simulated ranking for 34 countries, whilst there is less than or equal to five-rank difference for 152 countries (84.4 % of the cases). These results demonstrate that the CPI 2017 ranking is robust in the estimation of the 'global' parameters (mean and standard deviation) which are subsequently used to render the scores from the 13 sources comparable.

Table 6. *Differences in the ranking according to the method of imputation.*

Ranking changes	Percentage of countries
Less than or equal to 2 positions	50.0 %
Less than or equal to 5 positions	84.4 %
Greater than 5 positions	15.6 %
Greater than or equal to 10 positions	6.1 %
Greater than or equal to 20 positions (*)	0.6 %

(*) Comoros is the only country that changes more than 20 positions (specifically, 30 positions).

Source: Own elaboration.

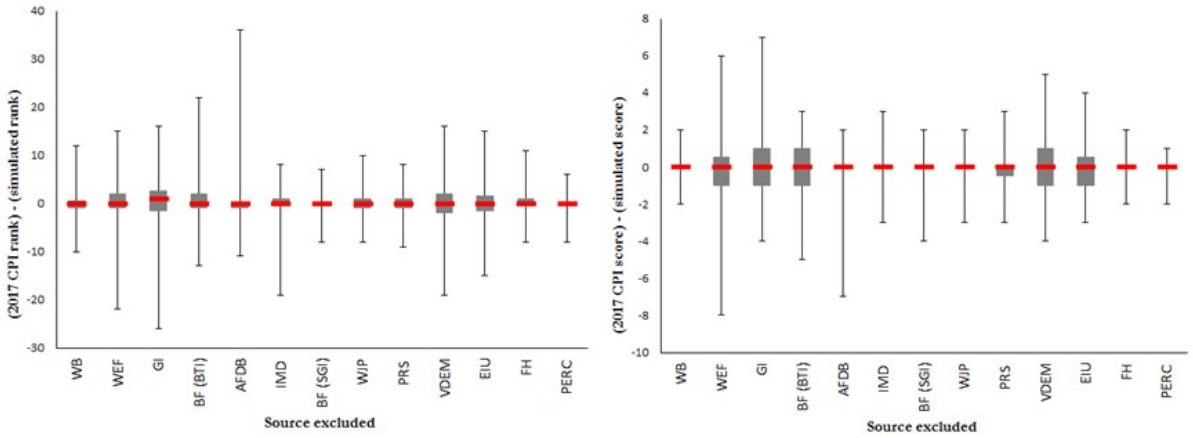
([7]) The EM algorithm is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating the following steps: (a) the expectation E-step: given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the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

Evaluating each source's contribution to the final CPI score

A further concern relates to whether the CPI is well balanced across the 13 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 ([8]). Instead, given that some sources have very limited country coverage, we tested the impact of each source on the CPI ranks and scores by excluding one source at a time. 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 sources (criterion for inclusion in the CPI).

Figure 3. *CPI framework: Impact on the CPI when one source is excluded at a time.*

(a) Impact on the CPI ranking (b) Impact on the CPI scores



Source: Own elaboration.

The main results of the impact on rankings and scores are provided in Figure 3. The red line is the median across all countries and the boxes include 50 per cent of the cases. The whole distribution of the differences is displayed by the vertical 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. Looking at Figure 3(a), the median rank is close to zero for all sources, and the box is within ± 2 positions. This 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 Development Bank Governance Ratings (AFDB), the World Economic Forum

([8]) 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 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 nonlinear aggregations; (d) it is not invasive, in that no changes are made to the index or to the correlation structure of the indicators.

Executive Opinion Survey (WEF), the Varieties of Democracy (V-DEM) and the Global Insight Country Risk Ratings (GI). Yet, the influence is moderate for the majority of the countries. In fact, the percentage of countries that experience a change less than or equal to four positions with respect to the CPI rank when excluding one source at a time, ranges from 72 % (V-DEM) to 96 % (PERC). Similar findings are observed for the impact on CPI scores (see Table 2(b)). In this case, the median of the score is close to zero and the box is within ± 1 points. In general, both the impact on scores and on ranks suggests that no source dominates the overall index in terms of ranking and score. Overall, all sources contribute to determining the CPI ranking, although some of them seem to have a higher impact than others (e.g. V-DEM vs. PERC). This finding is also corroborated by the analysis of reliability of the CPI 2017. The reliability, measured by the Cronbach's alpha value, is very high at 0.95 — well above the 0.7 threshold for a reliable aggregate. This means that the 13 sources are closely related as a group (i.e. they are internally consistent). The reliability of the CPI 2017 is not significantly altered when one source is excluded at a time (see Table 7). The Cronbach's alpha value wanders around 0.95, regardless of the source that is excluded. In order to support the fact that all sources contribute to determining the CPI ranking, the CPI 2017 is highly correlated with each one of the scores that results when one source is excluded at a time (see Table 7). In particular, the correlation ranges from a maximum of 0.95, when the source PERC is excluded, to a minimum of 0.81, when AFDB is excluded.

Table 7. *Cronbach's alpha and correlation analysis when one source is excluded at a time.*

		Cronbach's alpha	Correlation		
			Minimum	Maximum	Average
	CPI	0.95	0.81	0.95	0.90
Excluded source	**WB**	0.96	0.39	1.00	0.71
	WEF	0.94	0.39	0.95	0.71
	GI	0.94	0.44	0.95	0.80
	BF-BTI	0.95	0.37	0.96	0.75
	AFDB	0.96	0.44	0.82	0.66
	IMD	0.95	0.37	0.95	0.81
	BF-SGI	0.95	0.21	0.93	0.78
	WJP	0.94	0.61	0.96	0.82
	ICRG	0.94	0.66	0.94	0.83
	V-DEM	0.94	0.47	0.91	0.80
	EIU	0.94	0.43	0.93	0.79
	FH	0.95	0.43	0.96	0.80
	PERC	0.94	0.21	1.00	0.86

Source: Own elaboration.

6 Conclusions

The JRC analysis suggests that the Corruption Perceptions Index (CPI), besides being appealing for reasons of transparency and replicability, 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:

- After Global Insight Country Risk Service 2017, 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 viewpoints 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 policymakers 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 and related reading

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

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

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

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

List of figures

List of tables

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:
- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

N-NE

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

doi: 10.2760/974516
ISBN 978-92-79-96745-0

Publications Office