INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

IMPROVING THE CORRUPTION PERCEPTIONS INDEX: ADDITIONAL DATA SOURCES AND THEIR EFFECTS

Petr Jansky
Natalia Li

IES Working Paper 16/2021

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Jansky P. and Li N. (2021): "Improving the Corruption Perceptions Index: Additional Data Sources and Their Effects" IES Working Papers 16/2021. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

Improving the Corruption Perceptions Index: Additional Data Sources and Their Effects

Petr Jansky[a]
Natalia Li[b]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University Prague, Czech Republic
[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University Prague, Czech Republic, email (corresponding author): jansky.peta@gmail.com

May 2021

Abstract:
Corruption is notoriously hard to measure directly, and cross-country corruption indices thus often rely on indirect information such as country experts' or international businessmen's perceptions. Transparency International's Corruption Perception Index (CPI) is one such indicator that is often used by policy makers and researchers. The CPI is a composite index that, in its 2019 version, draws on 13 different data sources for calculation, with a threshold of at least three available sources for a country to qualify for a ranking. Until now, however, it has not been clear whether the data sources it uses are the only suitable ones. To assess this, we revisit the choice of these data sources and propose several improvements to the CPI methodology. Specifically, we identify up to five new data sources as potential candidates for inclusion. We examine the effects of including these additional data sources in two simulations: including all five data sources or only the four most suitable ones. Our results are mixed: the inclusion of new data sources would lower the standard errors of the CPI, but we identify a lack of correlation between the CPI and some of the data sources. We conclude by discussing trade-offs involved in including additional data sources in the CPI that may provide lessons for other composite policy indices.

JEL: C15, D73, O17
Keywords: Corruption; Perception-based indexes; Composite indexes; Data sources; Transparency International; Corruption Perception Index

Acknowledgements: We thank Roberto Kukutschka and Jon Vrushi for their helpful comments on an earlier version of this paper, as well as Jan Cvrček and Jan Zítek for their excellent research assistance. Petr Janský acknowledges support and data access from Transparency International for an earlier version of this paper. To ensure transparency and replicability, and in accordance with open science practices, the non-proprietary part of the database and code used can be found in the Open Science Foundation depository
(https://osf.io/2eyq5/?view_only=9f68f992286c437a854df52a4603f144).

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They should not be attributed to Transparency International or any other organisation. Any mistakes and all interpretations are the authors' own.

1 Introduction

Corruption is a worldwide phenomenon that is worth measuring despite the fact that both corruption and its perceptions are notoriously hard to measure. The most well-known indicator of corruption perceptions is the Transparency International's (TI) Corruption Perception Index (CPI). The CPI has been published annually since 1995 and its recent editions have covered around 180 countries. The CPI is a composite policy indicator calculated by averaging external data sources; in its 2019 edition there were 13 of these sources, which comprise expert assessments and surveys of business executives conducted by international and intergovernmental institutions, private sector firms and other organizations. Despite the breadth of these 13 sources, it is not clear whether they are exhaustive and whether the CPI could not be improved by adding new sources.

In this paper, we assess this by conducting a review of other sources which measure manifestations of public sector corruption that might potentially be included in the CPI. Our primary research question is what additional variables and data sources exist that measure public sector corruption and are not currently included in the CPI, and what effects their addition would have on the CPI. We use several approaches to review the existing data sources, ranging from internet searches to systematic review of academic papers. In addition to identifying additional data sources that are potentially suitable for inclusion in the CPI, we include a statistical analysis of the impact these additional sources would have on the CPI scores. We assess which of these data sources ought to be added to the current ones and what effects these additional data sources would have on the CPI.

To offer a preview of the results detailed below, we identify five potential new sources on the basis of literature review. Each of the data sources differs from the others in its key characteristics, such as frequency of publication, regions and countries covered; this means that the sources are suitable for inclusion in the CPI to varying degrees. In particular, we exclude one of these five potential new data sources, the World Bank's Business Enterprise Environment Survey, due to its irregularities and inconsistencies in terms of country coverage. We conclude that adding the remaining four identified new data sources would improve the estimation of the CPI and document the effects of adding them with a statistical analysis. Of course, these four sources also differ from one another in a number of aspects, which we discuss in detail below. For instance, two of the four identified sources are published only every three years, whereas the minimum frequency of the existing CPI sources has so far been two years; this could make the CPI rely on average on older sources. This represents a trade-off between being up to date and being more informative. The best approach to this trade-off depends on what the CPI is to be used for. For the purpose of observing changes over longer periods of time and for most of academic research, informativeness is likely more valuable, but for observing changes between two specific years or focusing on the latest edition of the CPI, it would be more important to have the most recent data possible, which is what the TI indeed does. These differences considered, we do recommend including the four additional sources, but we understand that for a variety of reasons TI might opt to newly adopt only some or none of the suggested sources.

With this paper, we contribute to a variety of relevant academic and policy literature, most of which centres around corruption measurement in one way or another. For example, Heywood & Rose (2014) argue that despite improvements there is still a relatively weak understanding of how best to measure corruption and how to develop effective guides to action from such measurement. Also, various alternative perceptions-based indexes of corruption exist; these were recently compared for the purpose of analysing the effects of corruption on institutional confidence (Pellegata & Memoli, 2016). Some corruption perception studies have focused on specific countries. For example, Gong & Wang (2013) use survey evidence to confirm a low tolerance for corruption in Hong Kong, whereas Song &

Cheng (2012) use an expert survey to document regional disparity in corruption perception across Chinese cities and Li et al. (2016) analyse the perception of anti-corruption efficacy in China. Also focusing on China, Ang (2020) critiques conventional bundled measures of corruption and presents an alternative that measures the prevalence of the four categories of corruption: petty theft, grand theft, speed money, and access money. Using a sample of citizens in Spain and Chile, León et al. (2013) show that response scales are used differently in these countries, which may lead to misleading conclusions when their corruption levels are compared. Moreover, relevant studies have been written by the authors of other composite policy indicators (Roodman, 2006) as well as by the European Commission's CPI composite policy indicators evaluators (Álvarez-Díaz et al., 2018). Last, but not least, the CPI has been used in numerous social science research papers, including those published in top economics journals; we draw on some of these in our methodology section.

The CPI has attracted substantial discussion and, indeed, one relevant strand of literature consists of papers that critically evaluate the CPI and discuss its disadvantages. For example, according to Warren & Laufer (2009), the CPI serves as a self-fulfilling prophecy that stimulates the status quo. This prophecy is reinforced, Warren & Laufer (2009) argue, by the fact that countries characterized by unfavourable corruption rankings are most affected by the addition of new countries because their corruption levels then seem to have decreased even if perceptions of corruption there have not in fact changed. Similarly, Shacklock & Galtung (2016) argue that neither the scores nor the rankings are comparable across years because the sources employed in constructing the CPI differ across countries as well as across years. Analogously, Mungiu-Pippidi & Dadašov (2016) point out that the CPI is not sensitive to change over short periods of time, while Rohwer (2009), argues that any comparison for a single country at two points of time must rely on the same set of sources. While expressing understanding for the CPI's launch in the 1990s, Cobham (2013) argues that the CPI only reflects a narrow, expert perception because its sources are insufficiently diverse and that its continued use is counterproductive when TI's Global Corruption Barometer and other alternatives are available. Using another alternative, Madichie (2005) proposes an interpretation of the CPI in conjunction with the Bribe Payers Index, which takes into account the criticism that the CPI is focused on countries where bribes are paid and does not target multinational corporations (Rose-Ackerman, 2007). While some of these disadvantages of the CPI are inherent and difficult to address, the diversity, coverage and quality of its input data can be improved in a relatively straightforward way. So, rather than producing another paper on the various disadvantages of the CPI, we propose and evaluate a straightforward, albeit partial, way in which TI could update the CPI methodology and thus address one of the downsides of the CPI in its existing form.

The rest of this paper is structured as follows. In Section 2 we briefly discuss our methodological approach to identifying additional data sources and testing what impact their inclusion in the CPI might have. We then present the main results in two sections: in Section 3, we summarise which additional sources we have identified and why; in Section 4, we present the outcomes of our statistical evaluation of potential additional data sources for the CPI. Section 5 concludes.

2 Methodology

In this section we briefly introduce the various methodological approaches we take in this paper. Note that when we refer to the CPI's methodology, we refer to its 2019 version, which is described in detail on TI's website (https://www.transparency.org/files/content/pages/2019_the CPI_Methodology.zip), including a full description of the 13 sources that were used for the 2019 CPI.

We identify additional data sources in two steps, which could be described as creating a long list and then narrowing it down to a short list. With the objective of identifying as many potentially suitable additional data sources as possible, we start by combining the following three approaches to locate additional data sources, directly or indirectly.

First, the CPI itself offers a range of potential additional data sources. In particular, a number of data sources were previously used for the CPI but are no longer, despite still existing and providing relevant estimates. We systematically review all the data sources used for the CPI since its inception in 1995 and investigate those that are still available in detail.

Second, we systematically review some of the best academic research in the field of economics focused on corruption. For this review, we build upon a recently released list of papers: a publicly available spreadsheet, updated by Gianmarco Daniele and Martino Gilli and, for the period 1995-1999 by Oasis Kodila, lists papers related to corruption - its causes, consequences and possible solutions - that were published in 15 general interest or top field journals in economics between 1995 and early 2019. There are 169 papers in this list and we review all of those to which we have secured access – 148 (i.e. all but 21 papers which require paid access).

Third, we search for additional data sources using a variety of Internet and academic search tools such as Google Scholar.

In the next step, we shortlist the identified data sources by following the same criteria that TI uses to select data sources for use in the CPI. First, the data source must be conceptually aligned, i.e. must cover a perception-based assessment of corruption in the public sector, based on expert evaluation and surveys; second, the methodology must be reliable and the source must originate from a reputable institution; third, the data must be quantitatively granular: the data must use at least a four-point scale of perceived corruption; fourth, the data must be comparable across countries and available for multiple years.

Having identified potential additional data sources via this two-step procedure (the results of which we report in Section 3), we then turn to a statistical procedure with which we evaluate the identified sources' impact on the CPI scores. We assess which of the identified data sources are most promising statistically, by comparing them with the existing CPI data sources and by simulating the impact of their potential inclusion in the CPI. We estimate the mean and standard deviation of the new sources for the same year as the data and add them to the baseline. We then estimate the 2018 CPI with 18 and 17 sources (with and without the World Bank's Enterprise Survey data), using this mixed baseline data. We present the results of the statistical evaluation, including principal component analysis, in Section 4.

3 Additional data sources for the CPI

In this section we present the additional data sources identified through the three approaches in the first step of our methodology as described above. First, TI changes the methodology it uses to calculate the CPI every few years, including new sources and removing old ones. The 13 data sources used in the 2019 the CPI calculation, together with the years of the data used, are as follows:

1. African Development Bank Country Policy and Institutional Assessment 2016
2. Bertelsmann Stiftung Sustainable Governance Indicators 2018
3. Bertelsmann Stiftung Transformation Index 2017-2018
4. Economist Intelligence Unit Country Risk Service 2018
5. Freedom House Nations in Transit 2018
6. Global Insight Business Conditions and Risk Indicators 2017
7. IMD World Competitiveness Center World Competitiveness Yearbook Executive Opinion Survey 2018

8. Political and Economic Risk Consultancy Asian Intelligence 2018
9. The PRS Group International Country Risk Guide 2018
10. World Bank Country Policy and Institutional Assessment 2017
11. World Economic Forum Executive Opinion Survey 2018
12. World Justice Project Rule of Law Index Expert Survey 2017-2018
13. Varieties of Democracy (V-Dem) 2018

For previous editions of the CPI, TI used a number of other sources, which are no longer in use for various reasons. One common reason for this is that a revision of the CPI methodology in 2012 simplified the aggregation of the different data sources and included just one year's data from each source. This was done to enable comparison of the scores from that time onwards; this had not been possible with some of the reports in the past. Consistency and the quality of the data (cross-country comparability, multi-year dataset) are the key aspects for selecting a data source to be included in the composite aggregation. Table A1 in the Appendix lists data sources that were previously, but are no longer used for the CPI.

Three of these 16 data sources still exist, despite no longer being used for the CPI; we describe their basic characteristics below. First, the Asian Development Bank Country Performance Assessment provided its latest information in 2018, when experts were asked to assess three dimensions of transparency, accountability and corruption in the public sector. This data source appears to be suitable for inclusion in the CPI. Second, the United Nations Economic Commission for Africa still publishes its African Governance Report every two years, but the report's contents have changed such that it is no longer suitable for CPI purposes. In its earlier editions, it included a governance indicator designed to measure corruption on the basis of a survey of national experts, which consisted of questions related to corruption in the judiciary, in legislature, at the executive level, and to tax collection and access to justice and government services. That indicator is no longer available, for example it is not included in the most recent, 2018 report, and therefore this not a suitable data source for inclusion in the CPI. Third, earlier editions of the World Bank's World Development Report included indicators that are only loosely related to corruption; these were used in versions of the CPI until 1998, but are no longer suitable for inclusion since the CPI methodology was revised.

Second, we carried out a review of economics papers focused on corruption. We systematically reviewed some of the best academic research in the field of economics focused on corruption using a list created by Gianmarco Daniele and Martino Gilli. The CPI was mentioned in 30 papers (20% of the 148 papers we reviewed). Figure A1 in the Appendix shows the number of papers mentioning the CPI in individual years, with the most papers in 2007, 2009 and 2010, but no clear trend over the years. Of these 30 papers, 14 made use of the results of the CPI, 13 referenced the CPI and 11 only stated in passing what it is that the CPI measures. Only four papers provided a critical view on the CPI and five introduced alternative corruption-related indexes or compared these with the CPI. The CPI plays an important role in nine of the reviewed papers.

Third, we used internet and academic journal search engines to identify relevant data sources. We searched for indicators of corruption that are similar to the CPI. One such alternative indicator is the Control of Corruption indicator, which is part of the World Bank's World Governance Indicators, and draws on 22 data sources. In addition, we studied critical scholarly discussions of the CPI in detail, which also revealed some additional data sources.

Altogether, these steps led us to identify five potential new data sources for the CPI, which we detail below. Overall, the first approach of the first step identified five potential new data sources, the second approach three of them (with one overlap) and the third approach none. During the shortlisting process (described above), we then ruled out two of the data sources from the second approach (so all except for the overlapping one). This left us looking at five sources, summarised in Table 1, that are potentially promising additional data sources for inclusion in the CPI. These five sources are: the Asian Development Bank Country Performance Assessments, the World Bank's Business Enterprise Environment Survey, the Global Integrity Index, the IFAD Rural Sector Performance Assessments and the Institutional Profiles Database. All five of these sources are used, along with 17 others, by the World Bank in the estimation of its Control of Corruption indicator within the World Governance Indicators. We provide a more detailed description of each source in the Appendix.

These five sources are publicly available and their data are based on surveys that are conducted regularly. The only source for which the data regularity is a potential issue is the Business Enterprise Environment Survey (BPS) by the World Bank. The surveys for the BPS provide good enterprise-level information on corruption perception, but they are conducted on a rolling basis with a span of approximately 3 years in each country, which means the country coverage varies each year. In the next section we estimate the CPI with the new proposed sources included, twice: first, including all five proposed sources, and, subsequently, excluding the BPS.

Table 1. Proposed data sources for the CPI

Data Source	Provider	Remarks	Availability	Country coverage	Year	Frequency
Asian Development Bank Country Performance Assessments (ASD)	Asian Development Bank	Transparency, accountability and corruption in the public sector. Indicators include 16 dimensions of policy and institutional performance. Responses are coded on a 6-point scale (1 through 6, good). CPA indicators are used by the Asian Development Bank to allocate concessional loans.	publicly available	25	2018	Annual
Business Enterprise Environment Survey (BPS)	World Bank	Enterprise level survey, includes questions on how often and how much firms make extra payments to get things done.	publicly available	143	2013-2015	Approximately every 3 years
Global Integrity Index (GII)	Global Integrity	Africa Integrity Index - perception of corruption, expert assessment	publicly available	54	2018	Annual
IFAD Rural Sector Performance Assessments (IFD)	International Fund for Agricultural Development	Accountability, transparency and corruption, expert assessment	publicly available	102	2018	Annual since 2004, every three years from 2015 onwards.
Institutional Profiles Database (IPD)	IPD	Includes corruption, expert assessment, latest year - 2016	publicly available	144	2016	Approximately every 3 years

Source: Authors.

Note: A more detailed description of these data sources is provided in the Appendix.

4 Statistical evaluation of potential additional data sources for the CPI

We now present the results of our statistical analysis of the impact of the additional sources on the CPI scores and on the coherence of the CPI as a composite indicator. We began by doing this for the five identified sources and then repeated the same analysis using only four new sources (excluding the BPS).

4.1 Adding five new sources to the CPI estimation

Using the CPI methodology, we added the five new indicators to the existing 13 sources in the 2018 CPI estimation. We estimated the mean and standard deviation of the new sources for the same year as the data and added them to the baseline. We then used this mixed baseline data to estimate the 2018 CPI from the 18 sources, which we summarise in Table 2. Adding the five new sources increases the number of sources used for the CPI estimation by 1.7 per country on average, and reduces the standard deviation of the CPI ranking. It also increases the maximum number of sources from 10 to 13.

Table 2. Descriptive statistics comparing original CPI and simulated CPI estimation (+ 5 sources)

Variable	No. of observations	Mean	Standard deviation	Min	Max
Original CPI 2018	180	43.17	18.96	9	87
Number of sources	180	7.07	1.79	3	10
Rank	180	88.98	51.64	1	180
Simulated CPI 2018	186	44.55	16.73	12	84
Number of sources	186	8.79	2.39	3	13
Rank	186	91.85	53.59	1	186

Source: Authors.

In the CPI methodology, the threshold for ranking countries is coverage by at least three sources. Remarkably, the addition of five new sources increases the number of countries observed from 180 to 186, adding Samoa, Micronesia, Tonga, Kiribati, Tuvalu, and Belize to the ranking. Table 3, below, shows that there were two existing sources covering these 6 countries and that two of our newly identified sources, IFAD and ADB, contributed to making up the minimum of three sources necessary for CPI ranking.

Table 3. Newly ranked countries in simulated CPI estimation.

Country	GI CRR	WB the CPIA	WJP	IFAD	ADB	The CPI estimate	No. of sources	Rank
Belize	2.33	-	0.41	0.7	-	47	3	71
Samoa	3.67	4	-	-	-0.5	46	3	72
Micronesia	4.33	3.5	-	-	-0.3	45	3	75
Kiribati	3.67	3.5	-	-	-0.3	41	3	88
Tonga	3	3.5	-	0.5	-0.5	40	4	96

Tuvalu	3	3.5	-	-	-0.4	37	3	115

Source: Authors.

4.2 Adding four new sources to the CPI estimation (excluding BPS)

We proceeded to estimate the CPI once again, using the same methodology, but this time excluding the World Bank Business Enterprise Environment Survey (BPS), for the reasons presented in the previous section relating to the survey regularity and country coverage in any given year. In this estimation we therefore used a total of 17 sources. After obtaining the results, we compared the CPI scores, ranking and number of sources obtained in the original CPI and both of our augmented CPI estimations – using five and four new sources respectively. Table 4 reports the results of that comparison.

Table 4. Descriptive statistics comparing original CPI estimation and simulated CPI estimations (+5 and +4 sources)

Variable	No. of observations	Mean	Standard deviation	Min	Max
Original CPI 2018	180	43.17	18.96	9	87
Number of sources	180	7.07	1.79	3	10
Rank	180	88.98	51.64	1	180
Simulated CPI (+5 sources)	186	44.55	16.73	12	84
Number of sources	186	8.79	2.39	3	13
Rank	186	91.85	53.59	1	186
Simulated CPI (+4 sources)	186	44.5	16.92	12	84
Number of sources	186	8.62	2.29	3	12
Rank	186	91.83	53.55	1	186

Source: Authors.

Naturally, in the estimation using 17 sources the maximum number of sources used in the final ranking is 12 rather than the 13 found in our estimation with 18 sources. However, the difference in the average number of sources per country is not much lower – 8.6 vs. 8.79. The number of observations is still the same with 17 sources as it was with 18 sources – 186 countries ranked. Table 5 demonstrates the results of the CPI scores and rankings obtained using the different sources – the original 13 sources, 18 sources (including five proposed), and 17 sources (including four proposed, without the BPS).

The addition of the new sources affects the CPI score results, bringing Singapore into first place as the country with the lowest perceived corruption level and moving Denmark and New Zealand down the ranking. Looking at the top three countries closely, Singapore's original CPI score was estimated using nine sources, while Denmark and New Zealand's scores both used eight. From our newly added sources, the IPD contains data for all three of these countries; it allocates all three countries identical scores of 1. The end results of our CPI estimations, however, were most likely altered by Singapore's high score in the Political and Economic Risk Consultancy Asian Intelligence ranking, where it is estimated to have the lowest perceived corruption in Asian region. Since that report is focused on Asia, Australia and the United States, it does not provide data for other regions.

Table 5. Comparison of CPI estimation results using different numbers of sources

country	Original CPI estimation			Estimation with 5 new sources			Estimation with 4 new sources (excl. BPS)		
	CPI score	Number of sources	Rank	CPI score	Number of sources	Rank	CPI score	Number of sources	Rank
Singapore	85	9	4	84	10	1	84	10	1
New Zealand	87	8	1	83	9	2	83	9	2
Denmark	87	8	1	83	9	2	83	9	2
Finland	86	8	3	80	9	4	80	9	4
Sweden	85	8	4	80	9	4	80	9	4
Norway	84	7	7	79	8	6	79	8	6
Netherlands	82	8	8	77	9	7	77	9	7
Switzerland	85	7	4	77	8	7	77	8	7
Luxembourg	80	7	9	76	8	9	76	8	9
Germany	80	8	9	76	9	9	76	9	9

Source: Authors.

The results of our simulations remain broadly consistent with the original the CPI estimation, with more sources added and a higher number of variables. Previous studies that assessed the statistical validity and coherence of the CPI, such as Álvarez-Díaz et al. (2018), confirmed that use of a higher number of sources for the estimation is associated with smaller standard errors, hence, we can expect increased precision in the corruption perception estimation from adding 1.7 sources per country on average. The results of our simulation exercise are indicative and subject to change depending on the baseline year (i.e. baseline year of mean and standard deviation for the new sources). Overall, we conclude that the newly added sources bring additional value to the CPI estimation and recommend that the four identified sources (without the BPS) should be added to the official calculation of the index.

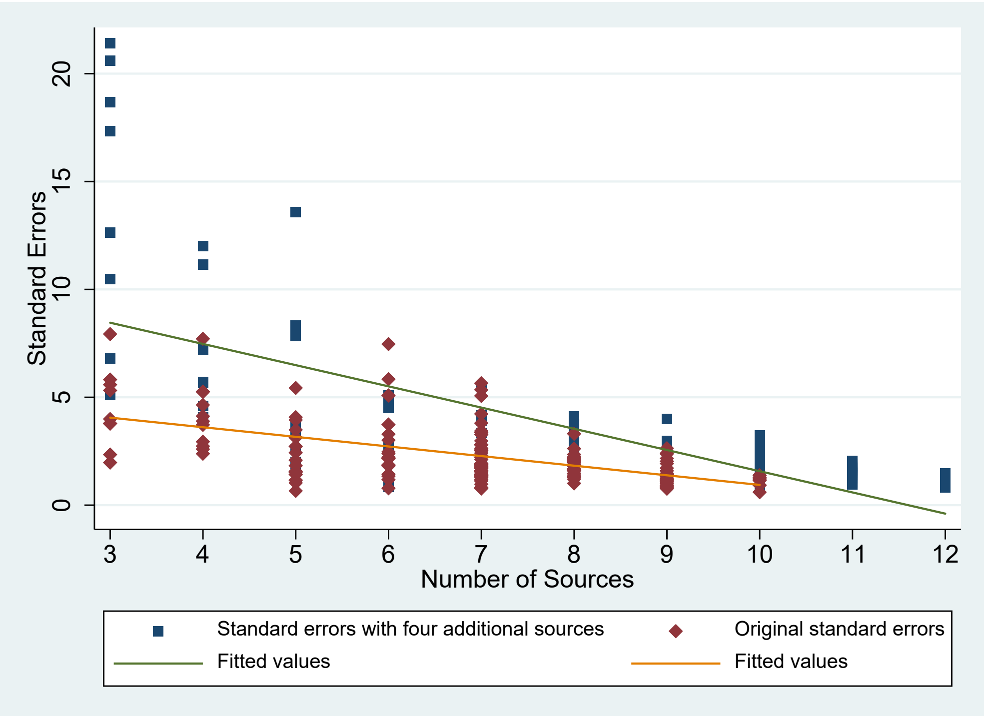
4.3 Assessing the impact of additional sources on the standard errors

We now assess the impact of the four newly added sources on the standard errors. A previous statistical evaluation by Álvarez-Díaz et al. (2018) suggested there is a negative correlation between the number of sources used for estimation and the standard errors of the CPI scores. Consequently, we expect the standard errors to decrease when we add the new sources. The general relationship trend as depicted in Figure 1 confirms that the negative relationship between the number of sources and the standard errors is similar for calculations using the original 13 and expanded 17 sources. As expected, this relationship continues to be negative for standard errors calculated with a greater number of sources and, on average, the standard errors are reduced when we add more sources to the calculation.

We then run a linear estimation of the correlation between the number of sources used to evaluate the CPI score of a country and the two different sets of standard errors – calculated with the newly added sources (17 sources) and without them (original 13 sources). Table 6 displays the results of two regressions estimating the effect of the additional sources on the standard errors. In regression (1), the dependent variable is *se2019* – standard errors associated with the CPI scores calculated with the original 13 sources, and the explanatory variable is the number of sources per country in the original CPI calculation *n_sources*. Regression (2) uses the new standard errors (*n_se2019*) as the dependent variable. The number of sources per country after the four new sources were included (*newbbps_sources*) is the explanatory variable in regression (2).

The results of regression (1) confirm the negative relationship between the number of sources per country used for calculating the CPI score and the standard errors. The estimated effect of the number of sources used per country on the standard errors is negative and significant at the 1% level. According to our estimation, adding one more source to a country's CPI score estimation reduces the standard errors on average by 0.45. The estimation results of regression (2) support the existing trend. As we add more sources, the negative effect of an additional source on the standard errors increases to 0.98 per country, on average.

Figure 1. Comparison of standard errors between original CPI calculation and CPI calculation with additional sources



Source: Authors.

The disadvantage of adding new sources and new countries to the CPI estimation is that it increases the variance of the standard errors for countries whose CPI score is estimated over three or four sources. This means an even greater bias for the countries for which only a small number of sources are available. In the CPI calculation with four additional sources, both the variance and the standard errors decline for countries whose CPI score is based on six or more sources. Our results reiterate previous studies' findings that, to reduce its bias, the CPI should increase the ranking threshold from three sources to five. In that case, the CPI would benefit from adding our four proposed additional sources, as doing so would increase the number of sources used per country, although the increased threshold would exclude 16 countries, based on our simulated estimation, leaving 170 countries in the ranking. The countries that would be excluded are mostly those with low- and middle- per capita incomes, including Solomon Islands, Tuvalu, St. Vincent and the Grenadines, St. Lucia, the Bahamas, Dominica, Barbados, Grenada, Brunei, Belize, Samoa, Micronesia, Suriname, Kiribati, Tonga, and

Maldives. Estimates for these countries based on a small number of sources, as in the CPI to date, may be biased or inaccurate, and this may have real political implications, especially in low-income countries.

Table 6. OLS estimation of the standard errors with different numbers of sources

VARIABLES	(1) se2019	(2) n_se2019
n_sources	-0.445***	-
	(0.0464)	
Constant	5.394***	-
	(0.338)	
newbbps_sources	-	-0.983***
		(0.0732)
Constant	-	11.41***
		(0.654)
Observations	180	186
R-squared	0.341	0.495

*Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1*
Sources: Authors

4.4 Assessing redundancy of information in the simulated CPI with added sources

In this section we assess the redundancy of the information in the CPI, including the newly added sources. Following the previous analysis by Álvarez-Díaz et al. (2018), we report Gamma statistics and Spearman rank correlations (Table 8). Similarly to Álvarez-Díaz et al. (2018), there is high correlation between the CPI and each of the sources, however, adding more sources reduces the correlation. In our analysis the correlation ranges between 0.4 and 0.9, compared to the previously reported range 0.87-0.95. Correlation between the sources is largely as expected and rather high; however, the correlation between the Africa Integrity Index (AGII) and the Institutional Profiles Database (IPD) appears to be unusually low. These two sources have 39 observed countries in common, but their correlation coefficients are 0.1 and 0.08 for the Spearman rank and Gamma statistics respectively. We suspect that the reason for this might lie in the methodologies that these particular sources use. The IPD is based on surveys that use four corruption-related questions about government-business and government-citizens relations, and assesses corruption perception in these areas. Meanwhile, the survey behind the AGII asks more specific questions on how corruption-related crimes are prosecuted and whether or not transparency mechanisms are enforced. As this is done via survey, the answers are also perception-based. Overall, these results are consistent with the earlier assessment by Álvarez-Díaz et al. (2018).

The proposed new sources appear to have relatively low correlation with the CPI score (from 0.4 to 0.85). Particularly, two of the new sources – the ADB and Africa Integrity Index (AGII) – have the lowest correlation with the CPI. We did expect relative changes in correlation, given that adding these sources to the CPI adds precision to the data for certain regions. Specifically, the ADB and AGII cover only countries from specific regions, with ADB focusing on Asia and the Pacific and AGII covering African countries. Our analysis thus confirms that the information used for the CPI estimation is not redundant. The new sources offer reduced correlation with the CPI score and focus on different numbers of countries, from 25 countries for the ADB to 144 countries in the case of the IPD. Also, the CPI provides greater country coverage when estimated using 17 sources, compared to the existing CPI estimated with 13 sources, as discussed in Section 4.

Last, but not least, we apply principal component analysis to assess redundancy of information in the simulated CPI with added sources. Adding our identified potential sources to the CPI estimation means that the CPI would have a set of seven sources that cover a shared set of 78 countries – WEF, GI, BF-BTI, PRS, VDEM, EIU and IPD, the last of which is a proposed additional source. We do not add the other three potential new sources to this analysis, since they cover fewer countries and, in some cases, do not have any countries in common with the other six sources. We apply the principal component analysis method to these seven sources (Table 7). In this case, the first two principal components explain around 82% of the total variance in the dataset. This is different from the principal component analysis on the six original CPI sources, where 80% of variance was covered by the first component. However, the seven sources still have comparable loadings on the first component, which supports the logic of giving these seven sources equal weighting in the CPI calculation. Thus, adding IPD to the principal component analysis supports previous findings that the CPI methodology of calculating an arithmetic average across the sources is statistically sound.

Table 7. Principal Component Analysis on seven sources

Principal Component	**Eigenvalue**	**Variance explained (cumulative)**	**Source**	**Loadings on the first principal component**
1	5.0	71.75	WEF	0.72
2	0.7	81.94	GI	0.86
3	0.3	86.80	BF_BTI	0.84
4	0.3	90.87	PRS	0.89
5	0.2	94.43	VDEM	0.86
6	0.2	97.47	EIU	0.91
7	0.2	100.00	IPD	0.84

Source: Authors

Table 8. Spearman rank correlations and Gamma statistics for 17 CPI sources (the original 13 + 4 new sources)

	THE CPI	ADB	IFAD	IPD	AGII	BF-BTI	BF-SGI	WEF	IMD	AFDB	GI	ICRG	WB	WJP	EIU	FH	PERC	V-DEM
CPI	-	0.40	0.69	0.68	0.51	0.73	0.81	0.57	0.72	0.72	0.85	0.85	0.85	0.74	0.93	0.75	0.69	0.74
ADB	0.40 (n=22)	-	0.53	0.07	-	0.41	-	0.76	-	-	0.32	1.00	0.43	0.43	1.00	1.00	-	0.27
IFAD	0.85 (n=101)	0.52 (n=14)	-	0.41	0.46	0.61	1.00	0.24	0.02	0.59	0.61	0.61	0.71	0.44	0.73	0.76	-	0.49
IPD	0.80 (n=144)	0.12 (n=10)	0.48 (n=79)	-	0.08	0.47	0.61	0.51	0.57	0.40	0.68	0.71	0.41	0.63	0.77	0.38	0.64	0.60
AGII	0.63 (n=54)	- (n=0)	0.59 (n=48)	0.10 (n=39)	-	0.08	0.48	0.05	-	0.47	0.38	0.41	0.50	0.10	0.26	0.26	-	0.33
BF-BTI	0.84 (n=137)	0.43 (n=14)	0.74 (n=92)	0.55 (n=115)	0.55 (n=50)	-	0.80	0.31	0.30	0.69	0.74	0.65	0.68	0.57	0.81	0.90	0.53	0.57
BF-SGI	0.87 (n=41)	- (n=0)	1.00 (n=2)	0.70 (n=41)	- (n=0)	0.83 (n=15)	-	0.56	0.67	-	0.84	0.78	-	0.81	0.75	0.95	0.33	0.78
WEF	0.75 (n=135)	0.67 (n=8)	0.33 (n=71)	0.67 (n=121)	0.06 (n=37)	0.40 (n=106)	0.64 (n=39)	-	0.77	0.26	-	-	0.18	-	-	0.15	0.69	-
IMD	0.89 (n=63)	- (n=1)	0.06 (n=14)	0.75 (n=63)	- (n=1)	0.39 (n=37)	0.78 (n=40)	0.92 (n=60)	-	-	0.72	0.72	-	0.72	0.77	0.43	0.72	0.61
AFDB	0.83 (n=37)	- (n=0)	0.71 (n=35)	0.45 (n=27)	0.57 (n=37)	0.76 (n=35)	- (n=0)	0.33 (n=23)	- (n=0)	-	0.48	0.56	0.72	0.55	0.67	-	-	0.43
GI	0.88 (n=186)	0.31 (n=22)	0.64 (n=101)	0.72 (n= 144)	0.41 (n=54)	0.75 (n=137)	0.87 (n=31)	0.69 (n=135)	0.78 (n=63)	0.48 (n=37)	-	0.86	0.83	0.80	0.91	0.83	0.81	0.76
ICRG	0.92 (n=140)	0.63 (n=5)	0.70 (n=73)	0.79 (n=124)	0.47 (n=37)	0.72 (n=109)	0.72 (n=41)	0.77 (n=117)	0.86 (n=63)	0.61 (n=25)	0.86 (n=140)	-	0.76	0.71	0.86	0.66	0.70	0.71
WB	0.86 (n=72)	0.40 (n=22)	0.74 (n=59)	0.39 (n=42)	0.51 (n=39)	0.64 (n=56)	- (n=0)	0.19 (n=36)	- (n=1)	0.71 (n=36)	0.75 (n=72)	0.68 (n=37)	-	0.70	0.67	0.71	-	0.61
WJP	0.88 (n=124)	0.40 (n=9)	0.59 (n=72)	0.75 (n=107)	0.12 (n=34)	0.69 (n=95)	0.87 (n=31)	0.78 (n=102)	0.89 (n=51)	0.71 (n=24)	0.84 (n=124)	0.80 (n=106)	0.70 (n=42)	-	0.82	0.53	0.80	0.68
EIU	0.93 (n=131)	0.68 (n=7)	0.69 (n=65)	0.77 (n=120)	0.27 (n=30)	0.77 (n=103)	0.72 (n=41)	0.73 (n=114)	0.83 (n=63)	0.62 (n=15)	0.86 (n=131)	0.85 (n=123)	0.46 (n=28)	0.84 (n=101)	-	0.83	0.81	0.84
FH	0.87 (n=29)	0.87 (n=3)	0.90 (n=9)	0.45 (n=26)	- (n=0)	0.94 (n=29)	0.92 (n=11)	0.18 (n=25)	0.53 (n=14)	- (n=0)	0.84 (n=29)	0.75 (n=20)	0.68 (n=5)	0.65 (n=20)	0.80 (n=23)	-	-	0.71
PERC	0.86 (n=15)	- (n=1)	0.06 (n=6)	0.82 (n=15)	- (n=0)	0.69 (n=11)	0.40 (n=4)	0.87 (n=14)	0.88 (n=13)	- (n=0)	0.88 (n=15)	0.79 (n=14)	- (n=1)	0.91 (n=14)	0.87 (n=15)	- (n=0)	-	0.77
V-DEM	0.90 (n=174)	0.30 (n=17)	0.66 (n=98)	0.73 (n=143)	0.45 (n=54)	0.70 (n=137)	0.84 (n=41)	0.64 (n=134)	0.80 (n=63)	0.53 (n=37)	0.82 (n=174)	0.82 (n=138)	0.63 (n=63)	0.84 (n=118)	0.86 (n=131)	0.85 (n=29)	0.89 (n=15)	-

Source: Authors

Notes: Low diagonal: Spearman rank correlation coefficients. Number of countries common to each pair of sources is denoted by n and reported in parentheses. Upper diagonal: Gamma statistic. All coefficients are positive, as all negative values were multiplied by -1 (for sources where lower scores represent less corruption.

5 Conclusion

The TI's CPI, which averages existing surveys of corruption perceptions, is likely the most influential cross-country indicator of corruption. While there are good reasons to be suspicious about the CPI, as its critics have pointed out, its prominence in media, policy and research make it an important subject of study. The CPI is a composite index that, in its 2018 version draws on 13 different data sources for calculation. In this paper, we revisited the choice of these data sources since it was not previously clear whether the data sources used in the CPI were the only suitable ones.

We made use of a variety of search tools to identify five additional relevant data sources, from both international development organisations and independent research institutes. We then simulated an estimation of the CPI with these five potential sources added, and a second such simulation including just four additional sources – we excluded the World Bank's Business Enterprise Environment Survey based on the irregularity of its country coverage. This simulation exercise increased the number of sources used per country on average by 1.6 and 1.7 after adding four and five new sources respectively, compared to the official CPI statistics. Both simulations, with four and five additional sources, enabled us to calculate CPI scores for six additional countries – Samoa, Micronesia, Tonga, Kiribati, Tuvalu, and Belize. Based on this simulation exercise and the World Bank's Business Enterprise Environment Survey's timing and consistency of country coverage, we do not recommend that the latter be used in future CPI calculations.

We proceeded to estimate the effect of the remaining four identified new sources on the standard errors, grouped by the number of sources used for a country's CPI score. The results display a greater bias for countries whose scores are based on a smaller number of sources (3-4), and smaller variance for countries whose scores draw on a higher number of sources (over 5). However, the results of the linear estimation show that adding new sources on average decreases the standard errors. Remarkably, the effect is greater in the estimation with 17 sources (including our four proposed sources) than in the regression with the original 13 sources. Furthermore, we made an assessment of potential information redundancy, which demonstrated that adding new sources does not disturb the methodological soundness of the CPI calculation and the information offered by our additional four sources is not redundant. Based on the simulation exercise and this statistical analysis, we propose that the following four sources be used in calculations for the CPI in the future: the Asian Development Bank Country Performance Assessments; the Global Integrity Index; the IFAD Rural Sector Performance Assessments; and the Institutional Profiles Database.

Additionally, our analysis shows that the correlation between standard errors and the number of sources decreases as the number of sources per country increases. Although raising the minimum number of sources required for a CPI score to be calculated from three to five would exclude 16 countries from the CPI, we recommend that TI consider doing so in order to reduce the bias for countries with small number of sources. Given the political importance and implications of the CPI, increasing this threshold might also lead to new and higher quality surveys being carried out in developing countries.

We conclude by discussing the general trade-offs involved in considering inclusion of additional data sources in the CPI. While the CPI is clear about what it measures, it is less clear about whether it aims to track the development of that phenomenon over time or to be the best measure of it at a given point in time. Clarifying whether it wants to be used for comparisons across countries or over time would help to guide its methodological choices, as it cannot suit both types of comparison. Perhaps a bigger, related question for TI and other users of the CPI is whether advocacy or research are their priority – an advocacy tool that needs to be launched every year will make different methodological choices than a consistent time series of corruption perception estimates designed to be used in cross-country regressions. In this paper we have highlighted some of these specific trade-offs – i.e. a high number of countries versus the lower standard errors that would result from increasing the minimum number of data sources for country inclusion from three to five – but we also leave a number of other relevant trade-offs to be settled by future research.

6 References

Álvarez-Díaz, M., Saisana, M., Montalto, V., & Moura, C. T. (2018). Corruption Perceptions Index 2017 Statistical Assessment. *JRC Mission*.

Ang, Y. Y. (2020). *China's Gilded Age: The Paradox of Economic Boom and Vast Corruption*. Cambridge University Press. https://www.cambridge.org/core/books/chinas-gilded-age/389BE063CCB6E75DDA144C36DABACD7A

Cobham, A. (2013). *Corrupting Perceptions. Why Transparency International's Flagship Corruption Index Falls Short*. http://www.foreignpolicy.com/articles/2013/07/22/corrupting_perceptions

Gong, T., & Wang, S. (2013). Indicators and Implications of Zero Tolerance of Corruption: The Case of Hong Kong. *Social Indicators Research*, *112*(3), 569–586. https://doi.org/10.1007/s11205-012-0071-3

Heywood, P. M., & Rose, J. (2014). "Close but no Cigar": the measurement of corruption. *Journal of Public Policy*, *34*(3), 507–529. https://doi.org/10.1017/S0143814X14000099

León, C. J., Araña, J. E., & de León, J. (2013). Correcting for Scale Perception Bias in Measuring Corruption: an Application to Chile and Spain. *Social Indicators Research*, *114*(3), 977–995. https://doi.org/10.1007/s11205-012-0185-7

Li, H., Gong, T., & Xiao, H. (2016). The Perception of Anti-corruption Efficacy in China: An Empirical Analysis. *Social Indicators Research*, *125*(3), 885–903. https://doi.org/10.1007/s11205-015-0859-z

Madichie, N. O. (2005). Corruption in Nigeria: how effective is the corruption perception index in highlighting the economic malaise? *World Review of Science, Technology and Sustainable Development*, *2*(3–4), 320–335.

Mungiu-Pippidi, A., & Dadašov, R. (2016). Measuring control of corruption by a new index of public integrity. *European Journal on Criminal Policy and Research*, *22*(3), 415–438.

Pellegata, A., & Memoli, V. (2016). Can corruption erode confidence in political institutions among European countries? Comparing the effects of different measures of perceived corruption. *Social Indicators Research*, *128*(1), 391–412.

Rohwer, A. (2009). Measuring corruption: a comparison between the transparency international's corruption perceptions index and the World Bank's worldwide governance indicators. *CESifo DICE Report*, *7*(3), 42–52.

Roodman, D. (2006). *Building and Running an Effective Policy Index: Lessons from the Commitment to Dvelopment Index*. Center for Global Development. http://www.cgdev.org/content/publications/detail/6661

Rose-Ackerman, S. (2007). Measuring private sector corruption. *U4 Brief*, *2007*(5).

Shacklock, A., & Galtung, F. (2016). Measuring the Immeasurable: Boundaries and Functions of (Macro) Corruption Indices. In *Measuring Corruption* (pp. 117–146). Routledge.

Song, X., & Cheng, W. (2012). Perception of Corruption in 36 Major Chinese Cities: Based on Survey of 1,642 Experts. *Social Indicators Research*, *109*(2), 211–221. https://doi.org/10.1007/s11205-011-9896-4

Warren, D. E., & Laufer, W. S. (2009). Are corruption indices a self-fulfilling prophecy? A social labeling perspective of corruption. *Journal of Business Ethics*, *88*(4), 841–849.

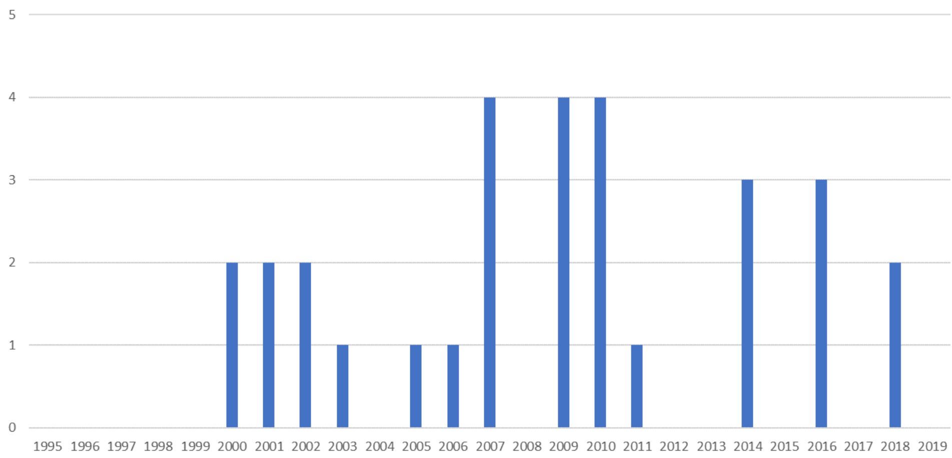
Appendix

Table A1. Data sources previously used for the CPI

1. TI Bribe Payers Survey – 2013
2. Asian Development Bank Country Performance Assessment – 2011
3. Merchant International Group, Grey Area Dynamics – 2008
4. United Nations Economic Commission for Africa, African Governance Report - 2007
5. CU, the State Capacity Survey by the Center for International Earth Science Information Network (CIESIN) at Columbia University – 2005
6. II, Information International, Beirut, Lebanon – 2005
7. WMRC, The World Markets Research Centre – 2005
8. TI/GI, Gallup International on behalf of TI – 2004
9. MDB, A multilateral development bank – 2004
10. PricewaterhouseCoopers, Opacity Index – 2003
11. The International Crime Victim Survey (ICVS) – 2000
12. Wall Street Journal, Central European Economic Review – 1999
13. World Development Report, World Bank – 1998
14. Internet Corruption Perception Index, Göttingen University – 1997
15. Impulse, Peter Neumann – 1996
16. Business International, survey, New York – 1995

Notes: The list shows the previous sources for the CPI, and the year denotes the last usage of the data. Source: Authors on the basis of TI's website.

Figure 1. Number of papers with the CPI



Source: Authors on the basis of a list created by Gianmarco Daniele and Martino Gilli (https://docs.google.com/spreadsheets/d/1E1LYkkWkU3OBiDW4B5ik_K2lSXAuMVaSA3O-VLu81oc/edit; accessed 30 November 2019).

Notes: Number of papers that mention the CPI out of the 148 reviewed papers published in top economics journals between 1995 and early 2019.

Description of five new data sources

1. **Asian Development Bank Country Performance Assessments**

The Asian Development Bank (ADB) is a regional multilateral development bank, established in 1966 with its headquarters in Manila, Philippines, and offices in 31 countries. Currently, the ADB has 68 members – 49 countries from Asia and the Pacific and 19 primarily developed countries from outside the target region. ADB provides loans, technical assistance, grants, and equity investments to its developing members and partner countries in order to reduce poverty in developing countries in Asia and the Pacific.

Country Performance Assessments (CPA) are conducted for all developing member-countries with access to concessional loans to ensure the effective use of development assistance. The composite country performance rating is calculated using the results of the expert surveys, which cover 16 criteria in 6 areas – economic management, structural policies, social inclusion and equity, policy and institutions, governance, and portfolio performance.

https://www.adb.org/sites/default/files/institutional-document/499546/country-performance-assessment-2018.pdf

Corruption assessment

Experts are asked to rate "transparency, accountability, and corruption in the public sector" on a scale from 1 to 6, where 1 is low and 6 is high level of transparency.

Country coverage

25 countries were scored in the 2018 assessment. As of 1 January 2018, 27 developing member countries had access to the Asian Development Fund (ADF). Sri Lanka and Viet Nam have been reclassified to group C effective from 1 January 2019 and are therefore no longer eligible for concessional assistance. Hence, country performance assessments (CPAs) were not conducted for those two countries.

Availability

The ADB has been conducting performance-based assessment since 2001, and since 2005 its reports on CPA rankings have been published on the bank's official website.

2. **Business Enterprise Environment Survey (BPS)**

The World Bank's Enterprise Analysis Unit conducts a firm-level survey of a representative sample of an economy's private sector. The surveys cover a broad range of business environment topics including access to finance, corruption, infrastructure, crime, competition, and performance measures. The topics covered in these Enterprise Surveys include infrastructure, trade, finance, regulations, taxes and business licensing, corruption, crime and informality, finance, innovation, labour, and perceptions about obstacles to doing business.

https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/Indicator-Descriptions.pdf

Corruption questions

The Enterprise Surveys include several indicators for corruption. Bribery Depth reflects the proportion of times a firm was asked or expected to pay a bribe when soliciting six different public services, permits or licences. Other indicators identify the extent to which specific regulatory and administrative officials require bribe payments during meetings with tax inspectors or to secure a government contracts. Another set of indicators focuses on bribes to obtain specific licences or permits, and shows the share of firms that are expected to make informal payments to secure import and operating licences and to obtain a construction permit

Country coverage

The surveys are conducted on a rolling basis, approximately once every 3 years in each country. A total of over 146,000 firms in 143 countries have been surveyed since 2005-2006. The dataset for this exercise contains information from 30 countries.

Availability

The dataset was retrieved from the World Bank World Governance Indicators portal for the years 2013-2015. However, in 2017 the World Bank stopped including the survey in WGI because the source is not annually updated. It can be accessed through the WGI portal http://info.worldbank.org/governance/wgi/#doc-sources

More public survey data is available on the Enterprise Surveys website in individual country datasets. This can be found here: https://www.enterprisesurveys.org/en/data

3. Global Integrity Index (GII)

Global Integrity is an independent, non-profit organization based in Washington, D.C., USA, tracking governance and corruption trends around the world using local teams of researchers and journalists to monitor openness and accountability. The Global Integrity Index uses around 300 indicators to assess the existence and effectiveness of anti-corruption mechanisms that promote public integrity. They typically pair an indication of the "in law" existence of a particular institutions with an "in practice" assessment of its functioning. The "in practice" assessment provides data on the perception of corruption and transparency.

Starting in 2013, the African Integrity Indicators also compiled by Global Integrity, for the 54 countries for which they are available. The AII follow a similar methodology and cover similar topics to the GII, but use a different structure of questions. We use averages of "in practice" questions from this source, corresponding to the categories shown in italics in the table below.

Corruption questions

1. In practice, allegations of corruption against senior level politicians and/or civil servants of any level are investigated by an independent body.
2. In practice, the body/bodies that investigate/s allegations of public sector corruption is/are effective.
3. In practice, appointments to the body/bodies that investigate/s allegations of public sector corruption support/s the independence of the body.
4. In practice, heads of state and government are investigated and prosecuted while in office if evidence suggests they committed a crime.
5. In practice, the mechanism for citizens to report police misconduct or abuse of force is effective.

The answers are on the scale from 0 to 100, where 0 is no transparency and accountability, and 100 is high level of transparency.

Country coverage

The 2018 dataset covered 54 African countries.

Availability

The data is public and can be accessed through the Global Integrity website. The data for 2018 can be found here https://www.globalintegrity.org/resource/africa-integrity-indicators-round-7-2019-xls/.

The World Governance Indicators also use the GII (AII), using a simple average of the "in practice" components of each of the indicated GII indicators.

4. **IFAD Rural Sector Performance Assessments**

The International Fund for Agricultural Development (IFAD) is an international financial institution and a specialized United Nations agency based in Rome, the UN's food and agriculture hub. IFAD's mission is to address poverty, increase food security and improve the nutrition in rural areas of developing countries.

The performance assessment is a corporate-level evaluation by IFAD country economists, who assess 12 dimensions of the rural policy environment on a 6-point scale. The assessments are used in IFAD's performance-based allocation system for distributing resources across countries.

https://www.ifad.org/documents/38714182/39711481/PBAS+CLE+-+Full+Report.pdf/15cb3af1-2e3f-43b2-a62d-8868734e23dd

Corruption questions

- Accountability, transparency and corruption in rural areas

Country coverage

The assessment covers 102 mostly developing countries partnering with IFAD.

Availability

The data is publicly available, although not all countries are covered every year. The data for 2018 was retrieved from the World Governance Indicators. It can be accessed through the WGI portal http://info.worldbank.org/governance/wgi/#doc-sources.

5. **Institutional Profiles Database (IPD)**

The Centre d'Études Prospectives et d'Informations Internationales (CEPII) is an independent French research thinktank based in Paris. It is part of the network coordinated by France Strategy, within the Prime Minister's services. The Institutional Profiles Database (IPD) provides an original measure of countries' institutional characteristics through composite indicators built from perception data. The database was designed in order to facilitate and stimulate research on the relationship between institutions, long-term economic growth and development.

The global database contains 127 indicators describing a broad range of institutional characteristics, structured in nine functions:

1) political institutions;

2) security, law and order, control of violence;

3) functioning of public administrations;

4) free operation of markets;

5) coordination of stakeholders, strategic vision and innovation;

6) security of transactions and contracts;

7) market regulation, social dialogue;

8) openness;

9) social cohesion and social mobility

http://www.cepii.fr/institutions/EN/ipd.asp

Corruption questions

IPD covers corruption in the following areas:

53 Level of "petty" corruption between citizens and the administration

54 Level of "political corruption" (e.g. vote buying, illegal campaign financing, bribery, etc.)

55 Level of corruption between public authorities and local businesses

56 Level of corruption between public authorities and foreign businesses

Respondents rate them from 0 to 4, where 0=high level of corruption and 4=very low level of corruption.

Country coverage

The 2016 edition of the database covers 144 developing countries.

Availability

The data is publicly available and is updated approximately every three years. The dataset can be accessed through the IPD website after the registration.

http://www.cepii.fr/institutions/EN/download.asp

The data can also be accessed through the World Governance Indicators website here Institutional Profiles Database (IPD).

The data on the level of corruption between government and foreign businesses is not available.

IES Working Paper Series

2021

1. Mahir Suleymanov: *Foreign Direct Investment in Emerging Markets: Evidence from Russia since the 2000s*
2. Lenka Nechvátalová: *Multi-Horizon Equity Returns Predictability via Machine Learning*
3. Milan Scasny, Matej Opatrny: *Elasticity of Marginal Utility of Consumption: The Equal-Sacrifice Approach Applied for the Czech Republic*
4. Javier Garcia-Bernardo, Petr Jansky and Vojtech Misak: *Common Agricultural Policy Beneficiaries: Evidence of Inequality from a New Data Set*
5. Petr Jakubik, Saida Teleu: *Suspension of Insurers' Dividends as a Response to the Covid-19 Crisis: Evidence from Equity Market*
6. Boris Fisera, Menbere Workie Tiruneh, David Hojdan: *Currency Depreciations in Emerging Economies: A Blessing or a Curse for External Debt Management?*
7. Vojtech Molnar: *Price Level Targeting with Imperfect Rationality: A Heuristic Approach*
8. Alex Cobham, Tommaso Faccio, Javier Garcia-Bernardo, Petr Jansky, Jeffery Kadet, Sol Picciotto: *A Practical Proposal to End Corporate Tax Abuse: METR, a Minimum Effective Tax Rate for Multinationals*
9. Evžen Kočenda, Ichiro Iwasaki: *Bank Survival Around the World: A Meta-Analytic Review*
10. Michal Kuchta: *Scenario Generation for IFRS9 Purposes using a Bayesian MS-VAR Model*
11. Jozef Barunik, Josef Kurka: *Frequency-Dependent Higher Moment Risks*
12. Petr Pleticha: *Who Benefits from Global Value Chain Participation? Does Functional Specialization Matter?*
13. Alex Cobham, Petr Jansky, Chris Jones, Yama Temouri: *An Evaluation of the Effects of the European Commission's Proposals for the Common Consolidated Corporate Tax Base*
14. Javier Garcia-Bernardo, Petr Jansky: *Profit Shifting of Multinational Corporations Worldwide*
15. Tomas Domonkos, Boris Fisera and Maria Siranova: *Income Inequality as Long-term Conditioning Factor of Monetary Transmission to Bank Interest Rates in EA Countries*
16. Petr Jansky, Natalia Li: *Improving the Corruption Perceptions Index: Additional Data Sources and Their Effects*

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz