

Lesné, Frédéric

Working Paper

Three Ways to Assess Corruption: How to compare indicators?

FERDI Working Paper, No. I16

Provided in Cooperation with:

Fondation pour les études et recherches sur le développement international (FERDI), Clermont-Ferrand

Suggested Citation: Lesné, Frédéric (2013) : Three Ways to Assess Corruption: How to compare indicators?, FERDI Working Paper, No. I16, Fondation pour les études et recherches sur le développement international (FERDI), Clermont-Ferrand

This Version is available at:

<https://hdl.handle.net/10419/269277>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Three Ways to Assess Corruption: How to compare indicators?

FRÉDÉRIC LESNÉ

➔ FRÉDÉRIC LESNÉ, Ph.D. candidate in Economics at Université d'Auvergne/CERDI since November 2012 specialises in issues of governance and corruption as well as service delivery in developing countries. // frederic.lesne@etu.udamail.fr

Introduction

By weakening the institutional frameworks through which economic agents interact, corruption constitutes a major obstacle to economic development. Because of its secretive nature, measuring this phenomenon has proven challenging. Numerous initiatives have nonetheless been launched in the past two decades with the objective to produce reliable measures of corruption. For a researcher, find a way through the “jungle” of corruption indicators is not always easy. Yet, a close look at the issue of corruption measurement is often an essential preliminary step to successfully apprehend corruption as an empirical research topic.

The aim of this review is precisely to evaluate the main categories of indicators that have been used more or less interchangeably from the middle of the 1990s for the study of macro-determinants and consequences of corruption, namely the indicators based on expert assessments, indicators constructed from surveys capturing the perceptions and experiences of corruption from households or businesspeople, and composite indicators of corruption. This study focuses on multi-year global and regional indicators that provide comparable scores across countries and time.

.../...

This document benefited from comments by Michaël Goujon, Patrick Guillaumont and Joël Cariolle. All remaining errors and omissions are mine..

LA FERDI EST UNE FONDATION RECONNUE D'UTILITÉ PUBLIQUE.
 ELLE MET EN ŒUVRE AVEC L'IDDRI L'INITIATIVE POUR LE DÉVELOPPEMENT ET LA GOUVERNANCE MONDIALE (IDGM).
 ELLE COORDONNE LE LABEL IDGM+ QUI L'ASSOCIE AU CERDI ET À L'IDDRI.
 CETTE PUBLICATION A BÉNÉFICÉ D'UNE AIDE DE L'ÉTAT FRANÇAIS GÉRÉE PAR L'ANR AU TITRE DU PROGRAMME « INVESTISSEMENTS D'AVENIR »
 PORTANT LA RÉFÉRENCE « ANR-10-LABX-14-01 ».

... /... The first three sections of this document are dedicated to the respective advantages and drawbacks of each category of indicators, making it possible to define the context in which those indicators are the most relevant. The fourth section provides a tool aiming at guiding researchers working on corruption in their choice of a suitable corruption measurement instrument by defining which indicators are the most appropriate depending on the objectives and specific constraints of their research question.

This document, along with the corresponding database of macro-indicators of corruption, is a contribution to the "Innovative Development Indicators" (IDI) programme of the Fondation pour les Études et Recherches sur le Développement International (FERDI).

1. Indicators constructed from expert assessments

Paulo Mauro's article on the impact of corruption on economic growth published in 1995 in *The Quarterly Journal of Economics* is among the first macroeconomic empirical studies on corruption. This article uses data from Business International (BI), which became soon after The Economist Intelligence Unit (EIU), a private institution producing decision-aiding tools for international investors, banks and multinationals. The version of the BI database used by Mauro in his article is composed of indicators of corruption, red tape and efficiency of the judicial system in 70 countries between 1980 and 1983. Risk factors were assessed by BI based on standardised questionnaires filled by their correspondents ("experts") based in the evaluated countries. Those questionnaires were collected and verified, first at the regional level and then again at the headquarters of the organisation in order to ensure the validity and consistency of the results.

1.1. Advantages of indicators constructed from expert assessments

Perceptions are useful

In his article, Mauro mentions a number of advantages and problems associated with the use of expert assessment data. In terms of advantages, he points out that *de-jure* indicators used in earlier studies on political instability can be misleading. An example he mentions is the number of changes of governments in a given period that characterises a number of countries otherwise democratically anchored. Mauro also notes that what really matters for investment decisions are the perceptions of investors of the level of risk, including the political risk, more than the actual level of those risks. The results of his article match this interpretation: the perception of the risk of corruption negatively affects economic growth mainly via its effect of investment decisions. More recently, Arndt and Oman (2006) confirm that, from an investor's perspective, perceptions matter on their own.

More generally, economic agents take some of their decisions based on their perceptions and experts are sometimes able to provide a direct evaluation of these perceptions. Soares (2004) showed that perception of a high level of corruption discourages victims of crime from reporting

cases to authorities. For Kaufmann and Kraay (2008), perception data are extremely useful even when they do not reflect the reality as economic agents operate largely on the basis of their beliefs.

For certain aspects of governance, analysts and experts are better placed to provide accurate information. For instance, according to Kaufmann and Kraay (2008), the very detailed questionnaire on national budgetary processes of the Open Budget Index (OBI) can only be filled by experts with an in-depth knowledge over those processes.

Scores are comparable: the World Bank's CPIA

A high level of comparability in expert assessments can be obtained more easily than with other types of corruption data. This is particularly the case when the same experts evaluate all countries. Many organisations producing this type of indicators set up a benchmarking system to enhance the comparability of scores attributed to different countries (Kaufmann and Kraay, 2008). Such is the case of the World Bank for their *Country Performance and Institutional Assessment* (CPIA).

The World Bank's CPIA is produced annually by the staff of the World Bank to assess the quality of institutional frameworks and policies of borrowing countries aiming at reducing poverty, ensuring sustainable growth and the effective use of development assistance. One of the 16 criteria forming the CPIA corresponds to the evaluation of *Transparency, Accountability and Corruption* (TAC) in the public sector. World Bank teams attribute a score ranging from 1 to 6 to each of those 16 criteria, with a higher score denoting a better performance. In order to ensure greater uniformity in the process of attribution of scores across countries, the World Bank initially evaluates a dozen of benchmark countries and provides its teams in charge of evaluations with detailed definitions corresponding to each score. Assigned scores are systematically reviewed before the finalisation of the indicator (Arndt and Oman, 2006). For Knack (2006), the CPIA is a hybrid indicator in the sense that expert-based ratings are partly centralised and partly decentralised. Scores originate from teams of experts based in the evaluated countries and regional offices. Those scores are then submitted to a centralised review process to enhance their comparability. In practice, however, the vast majority of scores attributed by decentralised units are unaltered. Since 2006 (corresponding to the CPIA 2005), scores for the 16 CPIA criteria are available for countries members of the International Development Organisation (IDA)¹. Prior to that date, evaluations were publicized only by grouping of countries into quintiles according to their level of performance (Arndt and Oman, 2006).

Relatively inexpensive data with wide coverage: the ICRG

Another advantage of expert ratings compared to other types of data on corruption is their relatively low cost of production. It is considerably cheaper to have an organised network of experts, such as the World Bank's country economists in the case of the CPIA, to fill in a

¹<http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/IDA/0,,contentMDK:21378540~menuPK:2626968~pagePK:51236175~piPK:437394~theSitePK:73154,00.html> [Accessed on 15 July 2013]

questionnaire than to develop and implement nationally representative household or business surveys in a large number of countries (Kaufmann and Kraay, 2008).

Corruption data compiled from expert assessments are in some cases available for a large number of countries over a significant period of time. This is the case for the *International Country Risk Guide* (ICRG) produced by The PRS Group, which has become very popular among researchers partly for its wide geographical (14 countries in 2013)² and temporal (since 1984) coverage.

The ICRG has been used as early as 1995 in another pioneer research paper from Knack and Keefer on the empirical relationship between institutions and economic performance. According to The PRS Group, the corruption component of the ICRG database evaluates political corruption, and more specifically the forms of political corruption that they consider to be the most damaging to foreign companies: excessive patronage, nepotism and suspiciously close ties between politics and business³. Like BI, The PRS Group charges users for access to their database. Another common feature between the indicators of corruption produced by the two institutions is the centralised nature of their development process. In the case of the ICRG, evaluations originate from a network of correspondents with a working knowledge of the situation on the ground but final scores are determined centrally by a small group of individuals (Knack, 2006).

Interest expressed by researchers during the mid-1990s for indicators constructed from expert assessment data, indicators often produced by firms specialised in risk assessment, can also be explained by the fact that few alternatives were available to them at that time (Treisman, 2007).

Those advantages explain to a large extent why indicators using expert ratings have been so popular and are still today very much used in the empirical macroeconomic literature on corruption. This method for measuring corruption has nonetheless several drawbacks.

1.2. Problems associated with indicators constructed from expert assessments

Perceptions do not necessarily reflect the reality of corruption

Perceptions of corruption do not necessarily reflect the reality of the phenomenon, although this criticism is not only valid for experts. Opinions may not be grounded on hard evidence and be biased. Consequently, cross-country differences in perceptions of corruption may reflect the influence of a number of factors unrelated to the actual level of corruption in those countries. Treisman (2007) mentions a number of such factors relevant for experts: a country's level of democracy, its level of inequality, the media coverage of corruption cases, anti-corruption governmental campaigns and the local culture of politically motivated accusations of corruption. Even when perceptions echo reality, a change in actual levels of corruption will affect perceptions only after a delay. Uncertainty over the duration of this time lag threatens empirical studies on corruption, especially those analysing panel data (Treisman, 2007).

² <http://www.prsgroup.com/icrg.aspx> [Accessed on 15 July 2013]

³ http://www.prsgroup.com/ICRG_Methodology.aspx#EconRiskRating [Accessed on 15 July 2013]

Those indicators are conceptually imprecise

Corruption indicators using perceptions from experts are often imprecise in what they intend to measure. In practice, it is frequent for researchers to apply indicators to measure concepts different from what these indicators were originally developed for. As Arndt and Oman note in their 2006 report on uses and misuses of governance indicators, the World Bank has warned researchers, including its own, against interpreting a CPIA score as a narrow measure of the quality of a country's public policies. Some analysts interpret the scores of Freedom House's *Freedom in The World* index, which evaluates various aspects of political rights and civil liberties in 195 countries and 14 territories (in 2012)⁴, including the extent to which corruption affects governments, as a proxy for the level of democracy when others infer from it a measure of governance or of the degree of respect for human rights (Arndt and Oman, 2006).

Those indicators are influenced by their audience and the agenda of their institution

The question whether experts adapt their ratings to their audience may legitimately be asked (Knack, 2006). According to Svensson (2005), indicators using opinions from experts generally target the private sector, and more specifically foreign investors, which lead them to measure corruption primarily from a business standpoint. For instance, the ICRG measures explicitly forms of political corruption that The PRS Group considers to be the most risky for its clients, mainly multinational companies. As Arndt and Oman (2006) note, big business interests are not necessarily those of society in general. This can introduce a bias in empirical studies using this type of indicators.

Kaufmann and Kraay (2008) do not find this criticism convincing. They observe that governance scores attributed by profit-making rating agencies whose main clients are private companies are highly correlated with the scores given by other types of institutions, including NGOs and public-sector stakeholders. For the two researchers, if such a bias existed in the assessment of corruption by risk rating agencies, the correlation found between their scores and those of other organisations immune to this bias would not be so high.

A similar criticism relates to biases supposedly resulting from agency relationships prevailing within institutions that produce indicators based on expert assessments. Knack (2006) mentions, among other indicators produced by international organisations, the case of the World Bank's CPIA which could be influenced by the interests of the country teams in charge of its development. For Knack, because the CPIA is of major importance for the determination of the amount of aid flows delivered by the World Bank, the teams based in the evaluated countries would directly benefit from attributing higher ratings than deserved. If they attribute unfavourable scores, country teams may also damage their working relationships with the governments in countries they evaluate. Knack (2006) notes, however, that a statistical analysis carried out on CPIA ratings brings no

⁴ http://www.freedomhouse.org/sites/default/files/FIW%202013%20Booklet%20-%20for%20Web_1.pdf [Accessed on 15 July 2013]

confirmation that scores attributed to IDA countries qualifying for financial support from the World Bank are overrated compared to other countries.

An ideological bias could also affect this category of indicators if the agenda of organisations in charge of their development influence their results. This bias would affect NGOs more particularly. Using business survey data, Kaufmann, Kraay and Mastruzzi (2004) showed that country corruption ratings produced by think tanks are not systematically related to the political orientation of the government in power in those countries, thus providing evidence against the existence of an ideological bias in the construction of indicators based on expert perceptions.

Experts are too similar (home bias)

A recurring criticism is that most experts who provide data for corruption indicators share a similar background. This could generate a bias in their ratings, referred to as “home bias”. Selection processes for analysts, rarely transparent in practice, would tend to favour certain types of experts, typically those originating from economically advanced countries. A direct consequence could be a systematic underestimation of the institutional quality in a number of countries due to an improper understanding of the local culture by those experts (Thompson and Shah, 2005). For Arndt and Oman (2006), the fact that experts share the same background exacerbates the risk of circular thinking as well as the distrust of developing countries’ governments towards expert assessments. However, according to Treisman (2007), it is unlikely that this effect is quantitatively significant. If it were substantial, the correlation found between corruption ratings by foreign and national experts would not be so important.

Experts are influenced by their emotions and prejudices (halo effect)

In his famous study on corruption and economic growth, Mauro (1995) mentions other disadvantages in using perception data from experts. The main problem he identifies is that experts may be influenced by the economic performance of the countries they assess. This bias, known in the literature as “halo effect”, is well documented (Roubaud and Razafindrakoto, 2010). It appears when countries with better economic performance receive, for that reason, better governance evaluations by experts (Kaufmann and Kraay, 2008).

Another concern is that judgments from national experts are affected by their feelings towards the government in power in their country (Kaufmann and Kraay, 2008). This could bias the levels and trends of indicators based on their evaluations.

Similarly, it is not implausible that indicators using expert opinions mirror their prejudices on the causes of corruption (Roubaud and Razafindrakoto, 2010). Institutional variations in a country could persuade experts to adjust their evaluations of the level of corruption, even without directly witnessing any change in corrupt behaviour. In this case, as Treisman (2007) points out, data reflect at least partially theories linking institutional framework and corruption that are supported by these experts.

Experts influence one another (herd effect)

Also problematic is the possibility that experts are influenced by the evaluation of other institutions when forming their own judgment, or use the same (imperfect) sources of information. For Knack (2006), changes in scores for indicators constructed from expert assessments do not always reflect the evaluation of a modification of actual conditions but very often a correction from previous years which, with hindsight, are found by those experts too low or too high relatively to other indicators produced by different institutions. This effect, commonly referred to as “herd effect”, undermines the quality of indicators using expert perceptions (Arndt and Oman, 2006).

Teams of experts, fragmented and changeable, have different opinions

One of the reasons explaining the popularity of indicators based on expert perceptions is that they allow meaningful comparisons of corruption scores across countries and time. Yet, with teams of experts changing regularly, the comparability of data in the time dimension is sometimes questionable. Similarly, if experts do not rate the same set of countries, cross-country comparisons can be dubious. For instance, Treisman (2007) note that The PRS Group does not clarify how they make sure that a given score attributed to a country in a specific year represents the same level of risk as the same score attributed to another country or year.

Evidently, experts can have divergent and even conflicting views on certain aspects of their evaluation. For Kaufmann and Kraay (2008), this should dissuade researchers to rely exclusively on a single evaluation. This issue is illustrated by the remaining gap between the CPIA scores of the World Bank and the African Development Bank, despite the recent harmonisation of their methodologies (Kaufmann and Kraay, 2008).

Lack of transparency

For Mauro (1995), it is not clear whether BI’s rating scale, ranging from 0 to 10, where 0 represents the highest level of risk, is everywhere consistent. Concretely, this means that the difference in terms of risk between two scores of 4 and 5 could be different from, for example, the gap between two scores of 7 and 8.

More generally, this type of indicators suffers from a lack of transparency in their methodology and clarity in their evaluation criteria. This is particularly problematic for the comparability of data and replicability of studies. It also generates further measurement error. According to Kaufmann and Kraay (2008), indicators using perceptions from experts are as useful as their evaluation criteria are properly documented. For the two researchers, the World Bank’s CPIA and the governance indicators from Freedom House are examples of indicators for which appropriate documentation is provided on the criteria used to attribute scores. They find regrettable that, on the contrary, many organisations producing such indicators refuse to publish their evaluation methodology. Knack (2006) is less definite about the level of transparency of the World Bank’s CPIA. Although confirming that the evaluation criteria as well as the reasons for the attribution of a specific score

are available in some detail, he notes that other aspects of the CPIA evaluation are more opaque, most notably the way weights are attributed to the different aspects of corruption listed in the evaluation criteria.

As Knack (2006) notes, when evaluation criteria are undisclosed, the lack of precision of the concept measured by broad and multi-dimensional indicators is exacerbated, as for the corruption indicators of EIU and the World Markets Research Centre (WMRC). The ICRG, a multi-dimensional indicator that has been used recurrently in the economic literature to measure corruption, is not very transparent either. Knack mentions that the ICRG has been recalibrated a number of times, causing unexplained “jumps” in scores attributed to countries from one month to the next. Treisman (2007) also noticed sudden and disconcerting changes in ICRG ratings that raise doubts about the quality of the evaluation. Amendments to the methodology and content of indicators can also make comparisons of corruption levels across time challenging (Knack, 2006), even though this problem is not specific to indicators constructed from expert perceptions. A criticism sometimes addressed to the World Bank’s CPIA is a lack of methodological transparency that threatens the comparability of the indicator over time (Arndt and Oman, 2006). The CPIA’s evaluation criteria, supposedly fixed and explicit, have been revised several times in the past (Knack, 2006).

Poor dissemination

In addition to a lack of transparency in their methodology, the outcomes of those indicators are not always freely available. This is obviously the case for commercial indicators produced by private firms which require users to pay a fee. Other indicators are simply not available to the vast majority of the scientific community. This radically hampers the replicability of studies using those indicators. Poor dissemination of corruption data constructed from expert assessments is an issue acknowledged by Kaufmann and Kraay (2008). They note that scores attributed to the 16 criteria of the World Bank’s CPIA, although partly accessible to the public since 2006 for IDA countries are still not available for all countries for which they are computed. Pre-2005 data have not been published, nor have been the historical data of the African and Asian Development Banks’ CPIAs. As for the *Freedom in The World* index produced by Freedom House, only the ranking of countries in three categories (“Free”, “Partly Free”, “Not Free”) is available. Specific aspects of the evaluation are unrevealed (Arndt and Oman, 2006).

Those indicators are difficult to interpret

Indicators using expert evaluations are rarely provided with a quantification of their measurement error. As Kaufmann and Kraay (2008) note, this situation has contributed to create a false sense of precision and unreasonable emphases on small differences in scores. The PRS Group, for example, does not report any margins of error for the ICRG (Arndt and Oman, 2006).

Those indicators are often characterised by an ordinal and finite scale, which makes their interpretation difficult. In practice, however, researchers usually treat them as cardinal measures (Svensson, 2005).

2. Indicators constructed from survey data

During the 1990s, an alternative to expert assessments for the evaluation of corruption started receiving growing attention: survey data on perceptions and experiences of corruption from a nationally representative sample of households or businesspeople.

While survey data address a number of drawbacks from expert evaluations, they do not entirely eliminate them and add their own limits.

2.1. Advantages of indicators constructed from survey data

They allow a more detailed evaluation of corruption : BEEPS and Enterprise Surveys

According to Knack (2006), the strength of nationally representative surveys of households or firms is in measuring the incidence of corrupt behaviour encountered by users of government services. Unlike indicators constructed from expert assessments, corruption indicators using survey data are not only able to measure the prevalence of certain forms of corruption but also their scale and the share of the population affected, based on direct experiences of corruption as reported by surveyed individuals or firms (Svensson, 2005). For Knack (2006), even though this approach focusses primarily on administrative corruption, it can also inform on some aspects of state capture, including the improper influence over laws and regulations affecting businesses. The BEEPS are an example of such surveys.

The *Business Environment and Enterprise Performance Surveys* (BEEPS) is the outcome of a joint project between the World Bank and the European Bank for Reconstruction and Development initiated in 1999 to evaluate constraints to private sector growth in transition countries. Since 1999, nearly 25,000 firm owners and managers have been surveyed in 29 European and Central Asian countries about the management of their company and their interactions with the state⁵. These surveys have made possible the development of indicators of corruption, state capture, lobbying and quality of the business environment that are comparable across countries. In addition, BEEPS data are sufficiently precise to link back those experiences and behaviours to firm and environmental characteristics (Knack, 2006).

In 2008, the BEEPS were restructured to make them fully compatible with the *Enterprise Surveys*, the centralised unit for firm-level data collection efforts within the World Bank. Those surveys, started in 2002 and covering 135 countries in 2013, mainly developing countries, cover a broad range of business environment topics, including corruption. The *Enterprise Surveys* are designed to be

⁵ <http://beeps.prognoz.com/beeps/MultiHandler.ashx?slc=AboutBeeps> [Accessed on 15 July 2013]

representative of the national private sector. The mode of data collection is face-to-face interviews with the owner or one of the top managers of the sample firms. According to the World Bank, over 90% of the questions “objectively” ascertain characteristics of the business environment, such as the number of days required to obtain a construction permit, or the need to pay a bribe to get such a permit. The remaining questions assess the survey respondents’ opinions on the obstacles to firm growth and performance in their country⁶.

Those indicators are more precise

Because businesspeople form a relatively homogenous group, it is possible to ask them more specific questions knowing that most, if not all of them, will be able to answer them meaningfully (Knack, 2006). Survey questions can therefore be designed in such the way that respondents and analysts do not need to rely much on their own interpretation of broad concepts to answer them. Instead of asking survey respondents whether they think “corruption is widespread”, a better approach could be, for example, to ask them whether they have been solicited for a bribe in the month preceding the survey (Kaufmann and Kraay, 2008). In order to increase response rates and the reliability of the information gathered, a practice now common for this type of surveys is to phrase questions indirectly to avoid implicating the respondent in corrupt behaviour (Svensson, 2003). This is the case for the BEEPS, but also for the World Bank’s *World Business Environment Survey* (WBES) which asked more than 10,000 managers in 80 countries between 1999 and 2000 if it was common for companies “in their line of business” to pay some irregular additional payments to “get things done” (Treisman, 2007).

Those indicators are less exposed to a number of biases

For Treisman (2007), data capturing direct experiences of corruption, although subject to limitations such as selective or imperfect memory and fear of retaliation, are less likely to pick up popular prejudices than the more subjective survey questions.

Because they rely on perceptions, corruption indicators capturing perceptions of households or businesspeople share certain limits with indicators based on expert perceptions. They have, however, some distinct advantages. In contrast to most expert assessments, surveys provide data that are largely independent from other evaluations of corruption. Most managers sampled in surveys like the BEEPS are unlikely to know the governance scores attributed by commercial risk-rating agencies or other groups of experts to the country in which they operate (Knack, 2006). Even if they were informed of those ratings, this would probably not affect their responses to a significant extent. More generally, most firm and household surveys are less likely to have errors that are correlated with other sources of data on corruption than, for example, assessments by risk assessment agencies (Kaufmann, 2008). For some surveys, however, this advantage is less clear. As Knack (2006) notes, the annual World Economic Forum’s (WEF) *Executive Opinion Survey* targets

⁶ <http://www.enterprisesurveys.org/Methodology> [Accessed on 15 July 2013]

specifically “business leaders” with international experience. It is more likely that those managers are exposed to international corruption rankings produced by other institutions. Additionally, the survey explicitly requests its respondents to evaluate their business environment with respect to global norms rather than reasoning in national terms. Knack (2006) mentions a third factor undermining the independence of the WEF’s *Executive Opinion Survey*: the fact that a significant number of partner organisations implementing this survey also work with the International Institute for Management Development (IMD) to implement a similar business survey (also named *Executive Opinion Survey*). Consequently, it is very likely that a number of businesspeople are sampled simultaneously in both surveys.

To some extent, WEF and IMD’s executive surveys are more similar to expert opinion surveys, where “experts” are managers of companies that are leaders in their respective industry. The majority of business surveys are less interested in the opinion of managers exposed to international markets than in the direct experience of corruption from local businesspeople. Moreover, WEF and IMD’s surveys only produce indicators of corruption at the national level, whereas most firm surveys (BEEPS, WBES...) are constructed so as to allow firm-level analyses, and collect to this end information on their characteristics (Knack, 2006).

Despite the fact that firm surveys are better developed (Knack, 2006), household surveys can also provide useful information regarding the state and evolution of corruption at the country level.

Household surveys : ICVS and GCB

By the end of the 1980s, a group of European criminologists followed a few years later by the United Nations Interregional Crime and Justice Research Institute (UNICRI) initiated the *International Crime Victims Survey* (ICVS) project. This survey is part of an international victimisation study analysing the patterns of a number of crimes, including the demand for bribes by government officials. The project’s initial objective was to collect criminal data that could be compared across countries. It became the responsibility of the United Nations Office on Drugs and Crime (UNODC) in 2003. Five rounds of surveys were carried out between 1989 and 2005, for a total of 140 surveys implemented in 78 countries⁷. Surveys used a combination of computer-assisted telephone surveys in developed countries and face-to-face interviews in developing countries. In most developing countries, survey samples were drawn from the capital or major cities, and therefore were not fully representative of the country’s population (Svensson, 2005).

First released in 2003, the *Global Corruption Barometer* (GCB) of the NGO Transparency International is a survey capturing citizens’ experiences and perceptions of corruption practices occurring in their country. The GCB was originally derived from a set of questions drawn from the annual survey *Voice of the People*, created in 2002 and implemented by Gallup International⁸. In countries not covered by Gallup International’s survey (19 countries out of 69 in 2009), Transparency

⁷ http://www.crimevictimsurvey.eu/About_ICVS_2010/History_and_new_pilot [Accessed on 15 July 2013]

⁸ http://www.transparency.org/research/gcb/gcb_2003 [Accessed on 15 July 2013]

International independently commission survey companies⁹. With the exception of a few countries¹⁰, the 2013 version of the GCB was implemented by the WIN/Gallup International network. The GCB survey sample is generally nationally representative, except in a limited number of countries where only urban areas are surveyed. Different approaches are used to administer the questionnaires: face-to-face, self-administration, telephone or, in more economically advanced countries, the Internet¹¹. The 2013 GCB update, the largest to date, surveyed over 114,000 people in 107 countries¹².

Firm vs. household surveys

For Knack (2006), there are fundamental differences between firm and household corruption surveys. First, the former tend to attribute more importance to experiences than to perceptions. Second, household surveys generally suffer from greater comparability problems than firm surveys such as the BEEPS. This problem is partly due to the fact that a number of household surveys, including the GCB and the ICVS, sometimes cover only urban areas when implemented in developing countries.

Beyond their differences, business and household corruption surveys also share a number of strengths. Margins of error can easily be computed for indicators constructed using survey data. This is the case for several BEEPS questions (Francisco-Javier Urrea, 2007). Another advantage is that they capture the views of the population. As Kaufmann and Kraay (2008) note, this is very useful as individuals decide on their future actions on the basis of their beliefs. Lastly, it is likely that official authorities in evaluated countries consider surveys of individuals and local firms a more trustworthy source of information than assessments from external experts, towards which they are generally sceptical (Kaufmann and Kraay, 2008).

2.2. Problems associated with indicators constructed from survey data

Data are not necessarily accurate

Because corruption is illegal, it is difficult to obtain reliable information from businesspeople and citizens on their true experiences with corruption. Survey respondents may deliberately underreport their exposure to corruption by fear of retaliation from official authorities. For Kaufmann and Kraay (2008), many reasons can explain why people provide inaccurate information to surveyors. A patient waiting in line in a hospital might erroneously think that people ahead in the line paid a bribe to get there. Kaufmann and Kraay also give the opposite example of an individual who pays a bribe and receives the expected benefit in exchange. This person, if satisfied with the transaction, will not consider themselves as a "victim" of corruption in this case.

⁹http://issuu.com/transparencyinternational/docs/global_corruption_barometer_2009_web?e=2496456/2192681
[Accessed on 15 July 2013]

¹⁰ http://www.transparency.org/files/content/pressrelease/GCB2013_FAQs_EN.pdf [Accessed on 15 July 2013]

¹¹ http://issuu.com/transparencyinternational/docs/global_corruption_barometer_2009_web?e=2496456/2192681
[Accessed on 15 July 2013]

¹² <http://gcb.transparency.org/gcb201011/> [Accessed on 15 July 2013]

If those measurement issues are not related in a systematic way to country characteristics, they do not represent a danger for cross-country data comparability (Svensson, 2005). Moreover, specific survey techniques can help improve the quality of the data collected. In this respect, the choice of the organisation implementing the survey in the field is crucial. Also, as mentioned earlier, survey questions can be formulated such as to avoid implicating respondents, while relying on the people using the personal experience to answer the questions. Corruption questions can be asked towards the end of the questionnaire, when surveyors had enough time to establish credibility and trust. Finally, corruption-related questions can be asked on different sections of the questionnaire to test the reliability of the data (Svensson, 2003).

The difficulty to obtain accurate information from surveys is particularly pronounced for certain forms of “grand” corruption, as those behaviours and activities (embezzlement, conflict of interest...) take place away from public’s scrutiny (Knack, 2006).

Those indicators are subject to specific biases

Alike experts, perceptions of surveyed citizens and businesspeople can be affected by factors that are not directly related to corruption, like the degree of ethnic heterogeneity of respondents’ communities (Olken, 2009) and the freedom of the press (Roca, 2010). In a study published in 2010, Kaplan and Pathania show that, paradoxically, firm owners’ perceptions of the business environment may be negatively correlated with the GDP growth rate. According to Knack (2006), perceptions of firm survey respondents on whether corruption is an obstacle to business are possibly affected by optimism and prevailing economic conditions. Individual characteristics such as education or age can also influence perceptions. Donchev and Ujhelyi (2013) therefore suggest that data obtained from the same survey implemented in two different countries are directly comparable only if the population of these two countries share the same characteristics along those relevant dimensions.

For Bertrand and Mullainathan (2001), one of the most acute problems brought by subjective questions is that perceptions may change, even over a short period of time. Attitudes may not even “exist” in a coherent form.

Respondents may be tempted to provide answers that do not make them look bad in front of the enumerator or force themselves to express a clear opinion on a topic they have never really thought about before the interview. People may also be wrong about their own attitudes. They can fail to reasonably predict their own behaviour or explain the reasons for their past actions (Bertrand and Mullainathan, 2001).

Cognitive factors can also undermine the quality of the information obtained through surveys. The ordering and formulation of questions matter, together with the scaling and ordering of possible answers (Bertrand and Mullainathan, 2001). Respondents may perceive corruption in absolute (the number of cases) rather than relative (the share of population affected) terms, which penalises large and densely populated countries. Respondents may also have a lower sensitivity to

corruption in regions where it is endemic, implying that estimations of the prevalence of corruption in low and high-corruption environments would not be equally accurate (Donchev and Ujhelyi, 2013).

A cultural bias threatens data comparability

The same questions asked in different countries can be diversely understood, especially when capturing perceptions. This can reduce the comparability of data across countries (Treisman, 2007). While doubting about its actual influence, Kaufmann and Kraay (2008) mention this cultural bias which originates from the fact that individuals living in different countries have contrasting norms about what constitutes corruption. Donchev and Ujhelyi (2013) confirm that populations may differ in essential cultural dimensions, e.g. whether it is acceptable to openly criticize one's current government.

Surveys require a certain degree of interpretation from respondents

Responses to survey questions are often measured on somewhat arbitrary scales (Kaufmann and Kraay, 2008). Knack (2006) illustrates this issue with a question taken from the WEF's *Executive Opinion Survey* that asks managers to assess the frequency of illegal payments in their line of business from seven possible answers ranging from "is common" to "never occurs", without providing any indication on how respondents should interpret this particular wording. Similarly, some questions may be vague and open to interpretation (Kaufmann and Kraay, 2008).

The sampling methodology is not always transparent

The fact that some organisations do not disclose their sampling methodology casts doubt on the true representativeness of their surveys. To illustrate this point, Kaufmann and Kraay (2008) refer to the WEF's *Executive Opinion Survey* which stated objective is to draw a survey sample that is representative of the distribution of firms in terms of sector of activity and size in each country. The WEF also indicates that firms are selected based on their size and scope of activities so as to ensure that executives benefit from international exposure, without clarifying how these two potentially conflicting objectives are reconciled. According to Knack (2006), the IMD provides even less information than the WEF on the size and composition of their *Executive Opinion Survey* sample.

Conducting nationally representative surveys of households in developing countries is particularly challenging (Knack, 2006). Even more serious is the fact that the composition of firm surveys can vary depending on the level of corruption. Some companies may be excluded from survey samples because they belong to the informal sector or choose deliberately not to exceed a certain size as a coping strategy against corruption. Similarly, individuals may give up on establishing their own business or end their activities given the high level of corruption in their country.

Data are costly and poorly disseminated

Poor dissemination of survey data limits their usefulness for research. Sometimes, only aggregate results are available, as for the 2010/2011 version of Transparency International's GCB¹³. The 2013 GCB data are, on the opposite, fully available on request¹⁴.

Survey development and implementation can be very expensive. The BEEPS, for example, were very costly to produce (Francisco-Javier Urra, 2007). This partly explains why so few corruption survey datasets are available to researchers (Svensson, 2005).

3. Composite indicators of corruption

In 1995, Johann Graf Lambsdorff, a young economist working for the NGO Transparency International, developed a composite index that combined several indicators of corruption derived from expert assessments. The initial objective was to strengthen empirical research on the causes and consequences of corruption by providing the research community with an index that could accurately compare levels of corruption across countries¹⁵. The 1995 index aggregates information gathered from seven surveys produced by three institutions (Business International, Political & Economic Risk Consultancy and the Institute for Management Development) and capturing opinions of experts, bank executives and business managers to produce an international ranking of perceived corruption affecting public administration and politics¹⁶.

Very soon, Transparency International's index, whose methodology evolved to become a few years later the *Corruption Perception Index (CPI)*, was enthusiastically received by the media and firmly denounced by the countries at the bottom of the ranking¹⁷. Many criticisms, more or less justified, also targeted the index's methodological groundings as well as its usefulness for corruption research.

The *World Governance Indicators (WGI)*, developed in 1999 by Daniel Kaufmann, Aart Kraay and Pablo Zoido-Lobaton at the World Bank are an attempt to respond to some of those criticisms. The *World Governance Indicators* are six composite indices measuring various aspects of governance at country level, including the prevalence of corruption with the *Control of Corruption (CoC)* index. This index captures the "perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests"¹⁸.

¹³ <http://gcb.transparency.org/gcb201011/results/> [Accessed on 15 July 2013]

¹⁴ http://www.transparency.org/gcb2013/in_detail [Accessed on 15 July 2013]

¹⁵ http://www.icgg.org/corruption.cpi_childhooddays.html [Accessed on 15 July 2013]

¹⁶ http://www.icgg.org/corruption.cpi_olderindices_1995.html [Accessed on 15 July 2013]

¹⁷ http://www.icgg.org/corruption.cpi_childhooddays.html [Accessed on 15 July 2013]

¹⁸ <http://info.worldbank.org/governance/wgi/faq.htm> [Accessed on 15 July 2013]

Like Transparency International's CPI, the WGI summarise opinions over the quality of governance from a large number of experts, businesspeople, but also citizens, surveyed by polling companies, think tanks, non-governmental organisations, international organisations and private firms¹⁹.

From the very beginning, the CoC has drawn on a larger pool of primary sources than the CPI. Until 2012, a condition for the inclusion of a data source into the CPI was that this source provided a ranking of countries and evaluated the prevalence of corruption rather than its impact. For Transparency International, the ICRG corruption indicators measure the political risk associated with corruption rather than corruption itself, which is why those indicators have not, until recently, been incorporated into the CPI (Svensson, 2005). The World Bank has decided not to make the same distinction and use the ICRG as a source of data for the WGI, including the CoC index.

Since 2012, the CPI adopts a wider definition and now includes not only primary sources that measure the extent of corruption but also some that assess the success of anti-corruption policies²⁰. In order to increase the statistical precision of their index, Transparency International decided early on to set a minimum of three data sources for a country to be included in the ranking, which is not the case for the World Bank's WGI (Treisman, 2007). The CPI underwent a fundamental methodological review in 2012. The index has been considerably simplified to improve clarity and better portray time trends in corruption perceptions²¹. Transparency International releases the CPI on a yearly basis since 1995. The country coverage of the index has considerably increased over time: from 41 countries in 1995 to 183 countries in 2011 and 176 in 2012. The World Bank's WGI were published once every two years from 1996 to 2002, and annually since 2002. In its latest update, the CoC ranks 212 countries and territories for the year 2011.

The CPI and the CoC have had – and continue to have – a considerable impact on research. Their success can be explained by the substantial benefits of using composite indices of corruption in empirical studies.

3.1. Advantages of composite indices

A wide geographical and temporal coverage

One of the advantages of composite corruption indices is their ability to track a larger number of countries over time than other types of indicators. As Knack (2006) notes, no individual source covers all countries and some indicators do not overlap at all in their country coverage. Combining data from different sources into a single index is a way to use information more effectively.

¹⁹ <http://info.worldbank.org/governance/wgi/resources.htm> [Accessed on 15 July 2013]

²⁰ http://cpi.transparency.org/cpi2012/in_detail/#myAnchor1 [Accessed on 15 July 2013]

²¹ http://cpi.transparency.org/cpi2012/in_detail/#myAnchor7 [Accessed on 15 July 2013]

Measurement errors and biases from individual sources can be attenuated

Another significant advantage of composite indices over other corruption indicators is their potential capacity to reduce the influence of measurement error affecting inherently every indicator. If measurement errors are largely independent from one individual data source to another, then errors tend to cancel out when data are aggregated (Knack, 2006).

For Kaufmann and Kraay (2008), it is useful to distinguish between two types of measurement error: a statistical error (e.g. the sampling error for firm and household surveys) and a measurement error associated with the fact that no single indicator can measure concepts as broad as governance or corruption on its own, regardless of how statistically accurate the indicator is. Aggregate indicators can usefully synthesize and provide an overview of a wide range of individual indicators. They can also help researchers exploit the complementarities between the different categories of indicators (expert assessments, firm and household surveys). Conversely, individual corruption indicators, or even a set of indicators taken from the same data source (e.g. BEEPS) may be defined too narrowly to answer certain questions (Knack, 2006).

Composite indicators can also attenuate the effect of source-specific biases. Data contained in the CPI and CoC are collected partly from domestic and international experts, but also from firm owners and managers and, in the case of the CoC, from a number of citizens of countries included in the index. The aggregation of multiple primary indicators into a single composite index reduces the impact of biases characterising individually each of those categories of corruption data.

Those indicators allow the explicit calculation of margins of error

Combining data sources enables researchers to provide a quantification of the measurement errors affecting the composite index and its primary sources (Kaufmann and Kraay, 2008). Margins of error can usefully remind users of corruption indicators that they cannot always interpret small differences in scores between countries and time periods. For Kaufmann and Kraay (2008), differences in scores can be interpreted only as long as their confidence intervals do not overlap. Following this rule, a substantial share of possible comparisons between the WGI scores of two countries, or two periods for the same country, appear to be statistically significant. According to Kaufmann and Kraay, this indicates that the WGI are particularly informative.

The success of composite indicators of corruption should not hide the many criticisms, with some targeting an indicator more specifically while others challenge the concept of composite indicators more generally.

3.2. Problems associated with composite indicators of corruption

A definition that is imprecise and uncertain

Among the CPI sources are assessments from local and international experts and surveys of businesspeople. A number of indicators included in the CoC evaluate the frequency of bribe

payments while others measure their size or the economic costs of corruption. Some sources describe political corruption and several others assess acts of corruption that take place at the administrative level. Indicators also differ in the extent of their geographical coverage (Treisman, 2007). Consequently, it is not easy to define precisely what these composite indicators intend to measure concretely. The combination of indicators measuring concepts that are somewhat distant from one another reduces the conceptual precision of the resulting composite index while favouring a reduction in measurement error, provided that measurement error is indeed a reason why indicator outcomes vary in the first place. Kaufmann and Kraay (2008) do not dispute the fact that aggregating individual data sources into a single composite index creates an inevitable loss of specificity.

For Knack (2006), the CPI and the CoC are not only imprecise in what they seek to measure but tend also to be relatively “uncertain” in their definition. The way those indices are calculated adds an element of uncertainty to the concept measured. Uncertainty is especially important when the rating criteria of primary sources are undisclosed and the aggregation method is opaque. In the case of the CPI (until 2012), primary sources are weighted equally, except for business surveys for which the last two years of data are included (which is no longer the case with the new methodology). Data sources constituting the CoC receive a weight that is proportional to their degree of correlation with the other primary sources, using a statistical method derived from the principal-component analysis. For Treisman (2007), the debate around which aggregation methodology is the most appropriate is not particularly relevant as those indices are ultimately highly correlated.

A definition that is unstable

In practice, composite indicators are implicitly defined by their primary sources and the way these sources are combined. As the number and type of data sources are likely to change over time and be different across countries, so does the implicit definition of corruption measured by composite indicators (Knack, 2006). This definitional instability creates issues of comparability for scores attributed to different countries and periods (Arndt and Oman, 2006).

Interpreting scores and their evolution is complex

Utmost caution is required when interpreting scores and their evolution. A modification of the type or number of primary sources in a composite index can alter scores, even though corruption (actual or perceived) remains unchanged (Knack, 2006). Kaufmann and Kraay (2008) argue that alterations in the database of primary sources only account for a minor share of changes in WGI scores over time. They also assert that the majority of cross-country comparisons of WGI scores are based on a substantial number of data sources in common, which should circumvent most comparability problems.

A similar issue arises when the methodology underpinning a composite index undergoes revisions (Arndt and Oman, 2006). Two major changes affecting the CPI were adopted in 2002: the exclusion

of surveys of the general public from possible data sources in favour of expert assessments and surveys of businesspeople and the adoption of a new procedure (“matching percentile”) to standardise the scores of individual indicators before they are averaged to form the index (Thompson and Shah, 2005). Thus, comparisons between pre- and post-2002 scores are particularly hazardous. Despite the major CPI revisions adopted in 2002 and 2012 and the frequent changes in its database of primary sources, no retrospective recalculation of the index has been performed by Transparency International. In contrast, the World Bank provides an update of past WGI scores every time their database of primary sources is revised²².

Until 2012, the CPI reused the same businesspeople surveys for two consecutive years, which had the direct consequence of reducing interannual variability in scores (Treisman, 2007). For this reason and others, teams working at Transparency International and the World Bank on the development of the CPI and the CoC have persistently warned users against interpreting small differences in scores from one year, or one country, to another. Because the number of countries included in those indices has changed over time (from 184 in 1996 to 212 in 2011 for the 2012 CoC update), comparing rankings across time-periods is not more appropriate. These problems have not prevented a number of researchers from analysing CPI data compiled as a time series cross-section panel (Treisman, 2007).

In the case of the World Bank’s CoC, the standardisation procedure used to rescale primary indicators may prevent users from tracking changes over time (Knack, 2006). For each WGI index and every year they have been computed, the average score of governance is set to 0 and the standard deviation to 1, by construction (Arndt and Oman, 2006). WGI indices, including the CoC, therefore cannot effectively isolate improving or deteriorating global trends in governance. According to Kaufmann and Kraay (2008), there is no clear evidence of a trend in any indicator constituting the WGI. The choice of a constant global governance level would then only translate an innocent choice of units that does not prevent comparisons of WGI scores over time. Arndt and Oman (2006) do not find this argument convincing as they believe empirical evidence of constant governance at the global level is also lacking. Furthermore, they note that for the WGI to be effectively comparable over time, not only the global governance level must be constant, but also its standard deviation (set to 1 by definition), which should not be considered as granted. Their third argument for the lack of comparability of WGI scores across time periods is that many primary sources do not allow meaningful comparisons of their scores over time.

Primary sources are not always accessible

For Knack (2006), if some primary indicators are constructed in a non-transparent manner, the resulting composite index is likely to be somewhat opaque as well, regardless of how transparent the aggregation procedure is. Limited access to primary sources reinforces opacity as users cannot replicate indices (Arndt and Oman, 2006). A common criticism made to the CPI, still valid despite

²² <http://info.worldbank.org/governance/wgi/resources.htm> [Accessed on 15 July 2013]

the recent changes in its methodology, is insufficient access to its primary sources. In opposition, the World Bank has been publishing the raw scores of WGI primary sources for several years.

Normalisation and aggregation techniques are complex and controversial

A “natural” method of combining individual indicators is to attribute them an equal weight in the composite index (Knack, 2006). The current version of the CPI weights each of its components equally. This was not exactly the case prior to the 2012 methodological revision as different rounds of the same business surveys were included as a separate source.

Before the 2012 revision, CPI primary sources were normalised using country rankings rather than raw scores before being aggregated. This method is questionable as it entails a non-negligible loss of information. In contrast, the weights of the WGI sources are determined by how closely related they are with others, using a statistical method derived from the principal-component analysis.

The hypothesis of independence of primary sources is doubtful

The objective of improving the accuracy of a composite index can justify attributing different weights to primary sources. Those weights should, however, be assigned on the basis on how informative these sources are, something that is not necessarily reflected in their cross-correlations (Knack, 2006). For the World Bank team in charge of the WGI, if sources are independent of each other, a source that agrees more with the others is by assumption less affected by measurement error and is therefore a more accurate measure of governance. This justification no longer applies if measurement errors are correlated across primary sources. Insofar as experts rely on the same (imperfect) sources of information to form their perceptions, it is conceivable that measurement errors in expert assessments are more correlated with one another than with measurement errors contained in surveys of citizens and businesspeople. In this case, the WGI aggregation method assigns excessively large weights to expert ratings, compared to other types of governance data.

If the aggregation procedure favours a specific type of data, biases associated with it can sensibly affect the outcomes of the composite indicator. According to Kaufmann and Kraay (2008), this issue is unlikely to be substantial in view of the lack of empirical evidence that perception errors are indeed correlated.

If measurement errors are correlated across data sources, the rationale for combining data into a single composite index is undermined. Kaufmann and Kraay (2008) acknowledge that aggregate indicators can only attenuate the component of measurement error that is truly independent across primary indicators.

Additionally, if errors are correlated, confidence intervals for composite indices are systematically underestimated (Svensson, 2005). For this reason, Knack (2006) believes that gains in statistical precision derived from aggregating corruption data into composite indicators are likely far more modest than claimed. Quantifying the degree of interdependence among primary sources is impossible. Nevertheless, according to Knack, if appropriate corrections were made to account for

the correlations in measurement errors across data sources, far less differences in WGI and CPI scores across countries and time would be significant.

4. Selecting the right corruption indicators

From a researcher's perspective, the decision to favour one indicator over another in an empirical study on corruption is likely not to be innocent. In fact, the results and conclusions of the study can fluctuate depending on which indicator is selected, as those indicators may not have the same coverage (leading to work on different samples) and are imperfectly correlated (when working on the same sample). There are no formal rules to follow for the choice of an indicator and very often this choice is made by the researcher on an *ad hoc* basis. The various categories of corruption indicators have different advantages and drawbacks, reviewed in detail in the first sections of this report. Knowledge about these pros and cons makes it possible to design a decision-making tool aiming at guiding researchers in selecting in a more methodical way one or more indicators suitable for their study depending on the objectives and constraints of their research question. This is the purpose of this section.

4.1. Does the choice of the indicator matter?

If all available corruption indicators produced very similar, if not identical, assessments of corruption, the issue of choosing among indicators would be immaterial. A close examination of rank correlations (Spearman) between several indicators reveals, on the contrary, that this choice does matter.

The two major corruption composite indices – the CPI and the CoC index – are highly correlated. The country rankings in the 2010 versions of the two indices exhibit a correlation close to 0.99. Hence, as Treisman (2007) notes, the debate around which of the two methods used for the CPI and the CoC is the most appropriate to normalise and aggregate primary sources is rather pointless. This close correlation is not particularly surprising as the two indices have a substantial number of primary sources in common.

The country rankings of corruption obtained with these two indices are also very much correlated with those of several indicators built from expert opinions (coloured in blue in the subsequent table on the next page).

Table of cross-correlations

	IPD (1)	IPD (2)	ICRG	GCI (1)	GCI (2)	GCB (1)	GCB (2)	WGI	CPI
IPD 2009 Petty corruption	1.000								
IPD 2009 Political corruption	0.821	1.000							
ICRG 2010 Corruption	0.718	0.656	1.000						
GCI 2010 Bribery	0.797	0.720	0.789	1.000					
GCI 2010 Grand corruption	0.693	0.648	0.786	0.902	1.000				
GCB 2010 Bribery	0.799	0.693	0.684	0.755	0.703	1.000			
GCB 2010 Corruption perceptions	0.551	0.542	0.653	0.656	0.727	0.569	1.000		
WGI 2010 Control of Corruption	0.866	0.840	0.830	0.922	0.843	0.801	0.628	1.000	
CPI 2010	0.855	0.814	0.825	0.947	0.869	0.803	0.632	0.985	1.000

Spearman correlations, all significant at conventional levels, are computed on a common sample of 70 countries and territories evaluated by the nine indicators included in the table.

The countries and territories are the following: Azerbaijan, Argentina, Australia, Austria, Bangladesh, Bolivia, Brazil, Bulgaria, Cameroon, Canada, Sri Lanka, Chile, China, Colombia, Czech Republic, Denmark, Ethiopia, Finland, France, Germany, Ghana, Greece, Hong Kong, Hungary, India, Indonesia, Ireland, Israel, Italy, Japan, Kenya, South Korea, Lebanon, Latvia, Lithuania, Malaysia, Mexico, Mongolia, Morocco, Mozambique, Netherlands, New Zealand, Nigeria, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Romania, Russia, Senegal, Serbia, Singapore, Vietnam, Slovenia, South Africa, Zimbabwe, Spain, Switzerland, Thailand, Turkey, Uganda, Ukraine, United Kingdom, Tanzania, United States, Venezuela, Zambia, and Taiwan.

The indicators examined are (from top to bottom):

- The evaluation of petty corruption (IPD 1) and political corruption (IPD 2) from the “*Institutional Profiles Database*” (IPD) developed in 2009 by researchers from the French Ministry for Economic Affairs and the French Development Agency.
- The “Corruption” component of the 2010 *ICRG* by the PRS Group.
- “Irregular Payments and Bribes” (GCI 1) and “Diversion of Public Funds” (GCI 2) indicators from the 2010 *Global Competitiveness Index* (GCI), produced by the World Economic Forum.
- “Percentage of people that have paid a bribe in the past 12 months” (GCB 1) and “Perceptions of Corruption in Public Institutions” (GCB 2) from Transparency International’s 2010/2011 *Global Corruption Barometer*.
- The “Control of Corruption Index” from the World Bank’s 2010 *World Governance Indicators* (2012 update).
- The 2010 version of Transparency International’s CPI.

For more information, refer to the FERDI database of corruption indicators (which includes neither the GCB nor the ICRG).

The networks of experts producing the ICRG and the IPD – both included in the CPI – generate corruption rankings very similar to the two composite indices, with correlations close to 0.8. The degree of correlation among those two indicators constructed from expert assessments is however significantly lower, not exceeding 0.7.

Composite indices are even more correlated to the corruption assessments performed by business managers in the framework of the World Economic Forum’s 2010 *Global Competitiveness Index* (GCI). The GCI’s *Executive Opinion Survey* is used in both the 2010 versions of the CPI and the CoC index,

which partly explains this strong correlation. As mentioned earlier, this annual *Executive Opinion Survey* targets specifically “business leaders” with international experience. As a consequence, this survey informing the GCI is probably more similar to expert opinion surveys, where “experts” are managers of companies leaders in their respective industry, than to surveys of local businesspeople. GCI indicators are nevertheless sensibly different from IPD and ICRG indicators, specifically in their evaluation of “grand” corruption.

Indicators obtained from the 2010/2011 GCB survey, which targets citizens, generate corruption rankings that are significantly different from the other categories of indicators, even though those indicators are still statistically correlated with the GCB indicators. This is particularly noticeable for the indicator of perceptions of corruption in public institutions, a phenomenon that the general population is likely to be less qualified to evaluate than other categories of respondents, particularly experts.

Unsurprisingly, indicators derived from the same sources (IPD, GCI or GCB) produce very similar rankings. The fact that the IPD, GCI and GCB indicators measure corruption at different scales allows us to demonstrate that the CPI and the CoC index are correlated more strongly with measures of administrative corruption than with indicators capturing elements of “grand” corruption.

Despite the fact that the country rankings of all indicators displayed in the table are positively correlated, the high variance in cross-correlations calls for caution. The choice of favouring one indicator over another in an empirical study is clearly nontrivial. The analysis of rank correlations shows that, with the exception of composite indices, indicators belonging to a same group (expert perceptions, experiences and perceptions of citizens and businesspeople, and composite indices) are not necessarily more correlated with one another than with other categories of indicators. Institution-specific methodological approaches could account for a substantial share of differences in rankings between indicators. For a researcher, the process of selecting a suitable indicator should therefore not be confined solely to deciding on the most relevant category of respondents but should also factor in a number of additional considerations, described in detail in the next section.

4.2. Criteria for choosing a corruption indicator

Criterion #1: the definition of corruption adopted

The *definition of corruption* adopted by the researcher for a study is the first element to take into consideration when selecting a corruption indicator.

There is no consensus on a universal definition of corruption (Svensson, 2005) and activities such as lobbying, private contributions to electoral campaigns or other types of influence (e.g. revolving door, gifts) may or may not, depending on the researcher’s views and interests, be included in the field of investigation. Some definitions of corruption are confined to the public sector, whereas

others encompass “private-to-private” misconducts such as collusions between private firms or abuse of power on the part of an employee or manager at the expense of their company. Regardless of the way corruption is defined in the study, this definition must be consistent with the one used for the construction of the selected indicator.

Some initiatives aiming at measuring corruption, including the ICVS, focus exclusively on interactions between citizens and public officials. Others intend to measure corruption occurring between firms, like the *Bribe Payers Index* (BPI) of Transparency International.

Very often, it is difficult to uncover the definition of corruption used by an institution for the construction of a corruption indicator. As for the subcategory *Accountability* of Mo Ibrahim foundation’s Ibrahim Index of African Governance (IIAG)²³, an explicit definition is rarely provided.

The definition of corruption mentioned by an institution producing a corruption indicator may sometimes differ substantially from its practical implementation. For example, Transparency International defines corruption as “the abuse of entrusted power for private gain”, applied to both the public and private sector. The CPI, however, only collects perceptions of corruption in the public sector. According to the World Bank, the CoC index captures “perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as “capture” of the state by elites and private interests”²⁴. It is worth noting, however, that the aggregation procedure arguably skews this definition in favour of some of those corruption forms by attributing them larger weights in the index.

The researcher must therefore carefully check that the definition of corruption used for the construction of a selected indicator corresponds not only in principle but also in practice to the way corruption is defined in the study.

Criterion #2: the researcher’s objective

The *objective* of the researcher has obviously an important role to play in the choice of an appropriate indicator.

A researcher can be interested in corruption perceptions more than in its actual prevalence, in which case they should clearly examine corruption indicators constructed from perceptions, either from experts, businesspeople or the general population. If the study requires an indication of progress regarding the fight against corruption, the researcher may look into indicators capturing citizens’ opinions about anti-corruption national policies, like Transparency International’s GCB, or one of the Bertelsmann foundation’s *Bertelsmann Transformation Index* (BTI) corruption indicators which assesses, based on evaluations by experts, the degree of governmental success to contain corruption.

²³ <http://www.moibrahimfoundation.org/downloads/2012-IIAG-methodology-EN.pdf> [Accessed on 15 July 2013]

²⁴ <http://info.worldbank.org/governance/wgi/faq.htm> [Accessed on 15 July 2013]

Likewise, a number of indicators are appropriate for time series analyses whereas others can only provide a “snapshot” of the state of corruption. Because of its specific features and the many revisions made to its methodology since 1995, pre-2012 CPI scores are not directly comparable over time, unlike other Transparency International’s global indices such as the GCB²⁵ and BPI²⁶.

Criterion #3: the type of corruption assessed

The selection process must also consider the *type of corruption* corresponding to the research question.

Composite indicators such as the CoC, CPI and IAG have a relatively large scope and therefore may be more relevant for studying corruption in broad terms.

If the study is about specific types of “petty” corruption (e.g. bribery, absenteeism of public officials), the researcher would be more successful exploring household and business surveys which are more able to provide reliable information on these forms of corruption than composite corruption indicators. In order to measure the prevalence of various forms of administrative corruption impacting businesses, the researcher could examine for instance the World Bank’s *Enterprise Surveys*²⁷. Alternatively, the *Rule of Law Index* produced by the World Justice Project provides indicators of the prevalence of three forms of corruption - bribery, extortion, and influence peddling - affecting the executive branch, the judiciary, the legislature, the police and the military in 97 countries, using a general population survey²⁸.

Expert assessments are usually the most reliable source of data to inform researchers about certain forms of “grand” corruption (e.g. embezzlement, conflicts of interests). Corruption indicators contained in the ICRG measure political corruption and more specifically the forms of political corruption that The PRS Group, the institution producing the ICRG, consider to be the most risky to foreign companies: excessive patronage, nepotism and suspiciously close ties between politics and business. The “Institutional Profiles Database” (IPD), developed by researchers from the French Ministry for Economic Affairs and the French Development Agency includes an evaluation of “grand” corruption between government agencies and private firms as well as a measure of the prevalence of political corruption (e.g. bribery to secure public office, rigged elections, vote buying). Data are obtained from a global survey involving experts based in the country offices of the two institutions²⁹.

Criterion #4: the appropriate measurement tool

The issue of measurement in relation to the research question addressed is also relevant for the choice of a suitable corruption indicator. Measuring the proportion of the population affected by

²⁵ http://www.transparency.org/files/content/pressrelease/GCB2013_FAQs_EN.pdf [Accessed on 15 July 2013]

²⁶ http://bpi.transparency.org/bpi2011/in_detail/#myAnchor7 [Accessed on 15 July 2013]

²⁷ <http://www.enterprisesurveys.org/Data/ExploreTopics/corruption> [Accessed on 15 July 2013]

²⁸ <http://worldjusticeproject.org/factors/absence-of-corruption> [Accessed on 15 July 2013]

²⁹ http://www.cepii.fr/institutions/doc/2007_02.pdf [Accessed on 15 July 2013]

corruption, the costs of corruption for society, the incidence of corruption activities or their scope require different sets of indicators.

For example, the corruption questions included in the ICVS provide data on the proportion of the population asked or expected to pay a bribe to government officials during the civil year preceding the survey as well as the rate of reporting to relevant authorities. *Afrobarometer*, a research instrument that assesses the social, political, and economic atmosphere in Africa measures the frequency of bribe payments by the population in the context of various interactions with government officials. This survey also captures citizens' perceptions about the involvement of public officials and politicians in corruption as well as their views on the way their government combats corruption.

Other indicators measure the quality of governance, beyond corruption. The CPIA *Transparency, Accountability and Corruption* (TAC) criterion, produced independently by the World Bank and the African Development Bank, evaluates not only state capture by private interests but also the extent to which the executive branch as well as public sector employees can be held accountable for their use of funds and the results of their actions and the access of civil society to information on public affairs³⁰.

The International Fund for Agricultural Development's (IFAD) *Accountability, Transparency and Corruption in rural areas* (ATC) indicator assesses the extent to which, at the local level, the government - at both the executive and legislative level - as well as public employees and elected officials can be held accountable to rural poor people for the resources used in the course of their actions³¹. Therefore, and in contradiction to what its name suggests, the ATC indicator does not measure corruption but rather the degree of accountability of public actors. The IFAD justifies this strategy by arguing that a high level of accountability discourages corrupt practices.

Similarly, one of the corruption indicators in Bertelsmann foundation's BTI measures the extent to which public office holders who abuse their positions are prosecuted or penalized. This indicator, along with the CPIA's TAC criterion of both the World Bank and the African Development Bank are present in the 2012 CPI update. According to Transparency International, an indicator is eligible to enter the new version of the index if it measures the effectiveness of corruption prevention as "this can be used as a proxy for the perceived level of corruption"³². It should be recalled that until 2012 Transparency International was excluding the ICRG from the CPI, arguing that the corruption indicators included in this index were not measuring corruption but the political risks associated with corruption. The ICRG is, however, involved in the new version of the CPI.

³⁰ <http://data.worldbank.org/indicator/IQ.CPA.TRAN.XQ> and <http://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/2012-CPIA%20Questionnaire.pdf> [Accessed on 15 July 2013]

³¹ <http://www.ifad.org/operations/pbas/docs.htm> [Accessed on 15 July 2013]

³² http://files.transparency.org/content/download/533/2213/file/2012_CPI_DataPackage.zip [Accessed on 15 July 2013]

The World Bank's CoC index also contains a number of indicators, notably the IFAD's ATC, which do not measure so much "perceptions of the extent to which public power is exercised for private gain"-what the index is supposed to capture - but rather the degree of accountability of political leaders and public officials.

Criterion #5: biases associated with indicators

Corruption indicators are affected by different biases, reviewed in detail in the first three sections of this document. A researcher must factor them in when searching for a suitable indicator as biases can threaten the quality of their study. Depending on the research objectives and constraints, some biases are more problematic than others. The fact that the different categories of corruption indicators (based on expert ratings, household and business surveys, and composite indicators) have different biases hence becomes an asset. If the researcher considers the cultural bias, the fact that individuals living in different parts of the world have different norms about what constitutes corruption, as the main threat to the quality of their study, indicators using expert opinions should be favoured as they are less subject to this particular bias than other types of corruption data, especially surveys of citizens and local businesspeople. Conversely, if the researcher is mostly worried about a "herd effect" that appears when opinions are formed using the same (imperfect) sources of information, indicators constructed from survey data on experiences of corruption become an appropriate alternative to indicators using expert ratings. It is also worth keeping in mind that composite corruption indicators have the potential capacity to reduce the impact of biases characterising individually each category of corruption data through the aggregation process.

Criterion #6: the temporal and geographical coverage

Most composite indices of corruption, including the CoC and the CPI, and a fair number of indicators using expert assessment data such as the ICRG cover a large number of countries over a significant period of time. This is one of the main reasons for their extensive use in empirical macroeconomic studies on corruption in the last two decades. Yet, most corruption indicators cover a more limited number of countries (e.g. Afrobarometer, World Bank's CPIA) and/or provide data less frequently (e.g. ICVS, WBES). Those indicators may nevertheless be relevant for cross-sectional or regional studies.

4.3. Lessons

Lastly, these are good practices worth reminding ourselves when working with corruption indicators:

- Identify and acquire mastery over the methodology used to construct the indicator
- Identify the genuine potential for comparability over time and/or between countries
- Select appropriate indicator(s) following the above-mentioned criteria.
- Acknowledge and document the limitations of the indicator and its potential biases

- Take measurement error seriously
- Test the robustness of the results with alternative indicators meeting above criteria
- Promote replicability by granting access to the data (whenever possible)

Conclusion

This critical review, along with the FERDI database of corruption indicators, provides a systematic analysis of the main initiatives aiming at measuring corruption at the macroeconomic level. The objective of this study is to determine the relative strengths and weaknesses as well as the context in which those indicators used in corruption research are the most relevant. This review is also a first step in the definition of directions for future research towards the improvement of current corruption indicators and the development of more effective measurement instruments. Findings from this research will in turn contribute to improving our understanding of corruption and back up the formulation of concrete policy recommendations to development actors.

Abbreviations

ATC	Accountability, Transparency and Corruption in rural areas
BEEPS	Business Environment and Enterprise Performance Surveys
BI	Business International
BPI	Bribe Payers Index
BTI	Bertelsmann Transformation Index
CoC	Control of Corruption
CPI	Corruption Perception Index
CPIA	Country Performance and Institutional Assessment
EIU	Economist Intelligence Unit
GCB	Global Corruption Barometer
ICRG	International Country Risk Guide
ICVS	International Crime Victims Survey
IDA	International Development Association
IFAD	International Fund for Agricultural Development
IIAG	Ibrahim Index of African Governance
IMD	International Institute for Management Development
IPD	Institutional Profiles Database
GDP	Gross Domestic Product
NGO	Non-Governmental Organisation
OBI	Open Budget Index
TAC	Transparency, Accountability and Corruption
UNICRI	United Nations Interregional Crime and Justice Research Institute
UNODC	United Nations Office on Drugs and Crime
WBES	World Business Environment Survey
WEF	World Economic Forum
WGI	World Governance Indicators
WMRC	World Markets Research Centre

References

- Arndt, C. and C. Oman. 2006. "Uses and Abuses of Governance Indicators." *Development Centre Studies*, OCDE.
- Bertrand, M. and S. Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey." *The American Economic Review Papers & Proceedings* 91(2): 67-72.
- Donchev, D. and G. Ujhelyi. 2013. "What Do Corruption Indices Measure?" University of Houston
- Kaplan, D. S. and P. Vikram. 2010. "What influences firms' perceptions?" *Journal of Comparative Economics* 38(4): 419-431.
- Kaufmann, D., A. Kraay and M. Mastruzzi. 2004. "Governance Matters III: Governance Indicators for 1996, 1998, 2000 and 2002." *World Bank Economic Review* 18(2): 253-287.
- Kaufmann, D. and A. Kraay. 2008. "Governance Indicators: Where Are We, Where Should We Be Going?" *World Bank Research Observer*, World Bank.
- Knack, S. and P. Keefer. 1995. "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures." *Economics and Politics* 7(3): 207-227.
- Knack, S. 2006. "Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators." *World Bank Policy Research Working Paper 3968*, World Bank.
- Mauro, P. 1995. "Corruption and Growth." *The Quarterly Journal of Economics* 110(3): 681-712.
- Olken, B. A. 2009. "Corruption perceptions vs. corruption reality." *Journal of Public Economics* 93(7-8): 950-964.
- Razafindrakoto, M. et F. Roubaud. 2010. "Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa." *World Development* 38(8): 1057-1069.
- Roca, T. 2010. "Assessing Corruption: Expert Surveys Versus Household Surveys, Filling the Gap." *Lare-Efi Working paper DT/160/2010*, Montesquieu-Bordeaux IV University.
- Shah, A. and T. Thompson. 2005. "Transparency International's Corruption Perceptions Index: Whose Perceptions Are They Anyway?" *Mimeo*, World Bank.
- Soares, R. 2004. "Crime Reporting as a Measure of Institutional Development." *Economic Development and Cultural Change* 52(4): 851-871.
- Svensson, J. 2003. "Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms." *The Quarterly Journal of Economics* 118(1): 207-230.
- Svensson, J. 2005. "Eight Questions about Corruption." *Journal of Economic Perspectives* 19(3): 19-42.
- Treisman, D. 2007. "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science* 10: 211-244.
- Urta, F.-J. 2007. "Assessing Corruption An analytical review of Corruption measurement and its problems: Perception, Error and Utility." Edmund A. Walsh School of Foreign Service, Georgetown University.



Créée en 2003, la **Fondation pour les études et recherches sur le développement international** vise à favoriser la compréhension du développement économique international et des politiques qui l'influencent.



Contact

www.ferdi.fr

contact@ferdi.fr

+33 (0)4 73 17 75 30