

Governance Matters III:
New Indicators for 1996-2002 and Addressing Methodological Challenges

Daniel Kaufmann and Aart Kraay¹

In the latest installment of our efforts to measure the quality of governance worldwide, we have constructed estimates of six dimensions of governance covering 199 countries and territories for four time periods: 1996, 1998, 2000, and 2002. These indicators are based on several hundred individual variables measuring perceptions of governance, drawn from 25 separate data sources constructed by 18 different organizations. We assign these individual measures of governance to six categories capturing key dimensions of governance (Voice and Accountability, Political Instability and Violence, Government Effectiveness, Regulatory Burden, Rule of Law, and Control of Corruption), and use an unobserved components model to construct six aggregate governance indicators in each of the four periods. The paper, the data, as well as a web-based graphical interface, are available at www.worldbank.org/wbi/governance/govdata2002.

As we have emphasized in our previous work, an attractive feature of the aggregation method we use is that it provides us with not only estimates of governance for each country, but also with measures of the precision or reliability of these estimates, for every country, indicator, and year. While the addition of data has improved the precision of our governance indicator relative to previous years, the margins of error associated with estimates of governance remain large relative to the units in which governance is measured. Figure 1 illustrates this using two of our indicators for 2002, Rule of Law, and Corruption. We rank all countries according to their estimates of governance, and report a vertical bar summarizing the range of statistically-likely values of governance for each country, with the mid-point of the bar representing our best estimate.

These substantial margins of error imply that cross-country comparisons of levels of governance based on this type of data should be made with due caution. This is particularly the case for changes over time, which in the vast majority of cases are small relative to the margins of error associated with our estimates of governance (particularly during a brief time span). Nevertheless, in those cases where changes over time in estimates of governance are large (for example, the recent deterioration of Rule of Law in Zimbabwe, or the worsening of Political Instability in Argentina), we generally find that there is broad consensus among our many sources as to the direction of change.

It is important to realize that these margins of error are a pervasive feature of cross-country governance (and other) indicators, and simply reflect the common-sense observation that individual sources ultimately provide a noisy signal of the unobserved "true" level of governance. It is therefore important that aggregate indicators combine the information from these imperfect sources as efficiently as possible, and moreover

¹ The authors are with the World Bank. The opinions expressed here are the authors' and do not necessarily reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

accurately represent the extent of the remaining measurement error. As we discuss in more detail in the [accompanying paper](#), we use a methodology which optimally weights each individual source according to its precision or reliability, and this results in substantial reductions in overall margins of error relative to more ‘naïve’ approaches which simply create unweighted average scores for each country. We also show that the methodology used to create margins of error matters. For example, in the full [paper](#) we show that the bootstrapping procedure used by Transparency International to construct margins of error for its Corruption Perceptions Index exaggerates the precision of this index, and this bias is particularly pronounced for countries with relatively few data sources.

In our latest paper we also address a number frequently-heard criticisms of subjective governance indicators:

1. Are subjective data useful for measuring governance, or should they be dismissed as “only capturing opinions instead of facts”.

The main reason we rely on subjective data is that, for many dimensions of governance, it is the only data that is even potentially informative. This is especially the case for corruption, which is almost impossible to measure directly given its illegal and clandestine nature. Nevertheless, interesting recent research has made an effort to document corruption “objectively”, for example by comparing differences in prices paid for similar objects in public procurement (although it should be noted that such differences may capture a combination of mismanagement or inefficiency factors as well as corruption). However given the immense data problems associated with such an exercise, it is unlikely that cross-country comparable objective measures based on this idea will be a viable alternative any time in the near future.² In contrast, we currently have a wealth of perceptions-based data on many dimensions of governance, from a growing number of cross-country surveys of firms and individuals, and also from think-tanks and commercial risk rating agencies. We interpret the broad agreement among these many disparate sources regarding cross-country differences in corruption as signaling that all of these subjective sources have useful information content. Furthermore, we do find that subjective data does contain significant ‘signal’ content, and that where there is a substantial perception content, such perceptions do matter in terms of the strategies and behavior of the economic agent. Finally, over time there have been significant improvements in questionnaire design, so that so called subjective variables increasingly rely on ‘experiential’ questions (rather than generic opinions), which are often quantified in a cardinal sense (for instance measuring percentage of revenues paid as bribes, etc.).

² See for example Golden, Miriam and Lucio Picci (2003). “Proposal for a New Measure of Corruption, and Tests Using Italian Data”. Manuscript, University of Bologna, and DiTella, Rafael and Ernesto Shargrotsky (2003). “The Role of Wages and Auditing During a Crackdown on Corruption in the City of Buenos Aires”. *Journal of Law and Economics*. April.

2. Since the margins of error of subjective governance indicators are large, shouldn't we rely on "objective" indicators that do not have these measurement problems?

It is important to emphasize that objective measures of governance also have measurement error, and hence should also have associated margins of error as well.³ Consider for example using the share of trade tax revenue in total revenues to capture the inability of a government to broaden its tax base. This will be a "noisy" indicator of overall government effectiveness for at least two reasons: (i) the tax revenue itself may contain a variety of errors, and (ii) the extent of the tax base is only one of many dimensions of government effectiveness. The same is true for many other objective indicators of governance that have been proposed. Our calculations in the [paper](#) suggest that measurement error in these many objective sources is at least as important as the measurement error in subjective governance indicators: we impute margins of error for a number of objective governance indicators and find that they are generally as large as or larger than those associated with our indicators. This implies that problems of imprecision are likely to be pervasive regardless of the type of data used to measure cross-country differences in institutional quality.

3. Don't the subjective perceptions of think tanks and commercial risk rating agencies simply reflect the ideological biases of these institutions?

While this criticism has some visceral appeal, we find that it has surprisingly little substantive merit. In order to isolate the effects of potential ideological biases in governance ratings produced by these types of organizations, we compare their ratings with the responses of firms and/or individuals in cross-country surveys. Since the latter type of data will not reflect any ideological biases (since the respondents form a very large sample of individuals), systematic differences between the two types of ratings should be large if ideology plays an important role in the responses of think tanks and commercial risk rating agencies. In particular, we examine whether these organizations tend to give systematically higher or lower scores than surveys to countries with left-wing or right-wing governments. However, we find this is not the case. For the vast majority of think tanks and rating agencies, we find no evidence at all of systematic ideological biases, and in the one case where we do find some evidence, the effects tend to be small on average -- with the difference in ratings of countries with left- and right-wing governments of only about 10 percentile points. All this implies that on balance when utilizing all sources from think tanks and rating agencies there is no ideological bias.

4. Since these governance indicators only capture countries' relative positions, isn't it possible that some countries' relative governance scores worsen simply because the rest of the world is getting better?

It is the case that our governance indicators only rank countries relative to each other. We do this because we have relatively little information on absolute trends over time in governance. However, the limited information that we do have if anything suggests a small worsening in governance worldwide. We have reviewed the global

³ See Knack, Steve and Mark Kugler (2002). "Constructing and Index of Objective Indicators of Good Governance". Manuscript, The World Bank. for a variety of proposed objective governance indicators.

averages of the individual sources that are available in a consistent format over time, focusing on a common set of countries available since 1996. While most of the changes in these global averages are small, most of the statistically significant ones point to deteriorations in governance. Whether this reflects a true worsening of institutional quality, or reflects other factors at play, is an open question. However, it certainly is not the case that there is evidence of a worldwide improvement in governance, and thus downward trends in governance in individual countries cannot be justified with the argument that the world at large is getting better, and therefore they may be getting left behind even without a deterioration in absolute terms.

5. *Are these governance indicators sufficiently informative to be used as a basis for aid allocation or other policy decisions?*

This question has become relevant with the recent proposal of the U.S. government to rely in part on our governance indicators to allocate funds from the new Millennium Challenge Account ([MCA](#)). In order to be eligible, low income countries need to score well on a number of indicators of good governance, including several of ours. Most prominently, countries need to be in the top half of low-income countries on our Control of Corruption indicator in order to be eligible. Given the compelling evidence that aid works better in a good institutional and policy environment, explicit proposals such as these to allocate aid to countries with good institutions and policies are very welcome and encouraging. However, it is important to keep in mind that simply classifying countries as “in or out” according to a mechanistic rule such as this risks misclassifying countries, given the substantial margins of error present in the governance indicators. We show this in Figure 2, which shows Control of Corruption for countries potentially eligible for the MCA. While for countries well below and above the median, the risk of misclassification is low, for countries near the median there is a non-trivial risk that these countries will be misclassified by a simple “in or out” rule. This points to the importance of complementing the information in cross-country indicators such as ours with more nuanced and in-depth information from country governance diagnostics.

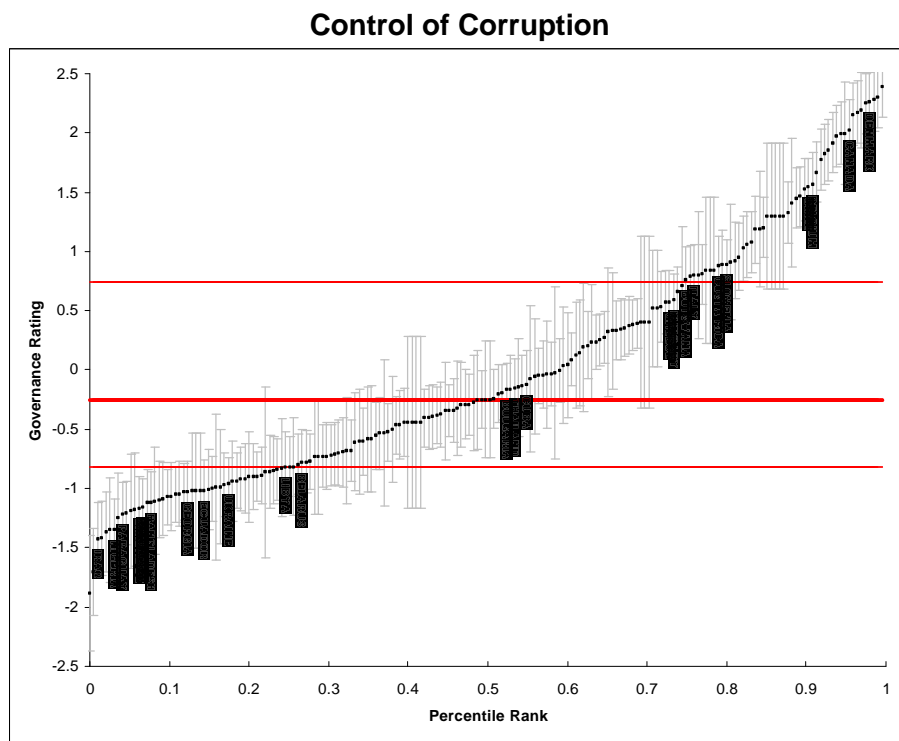
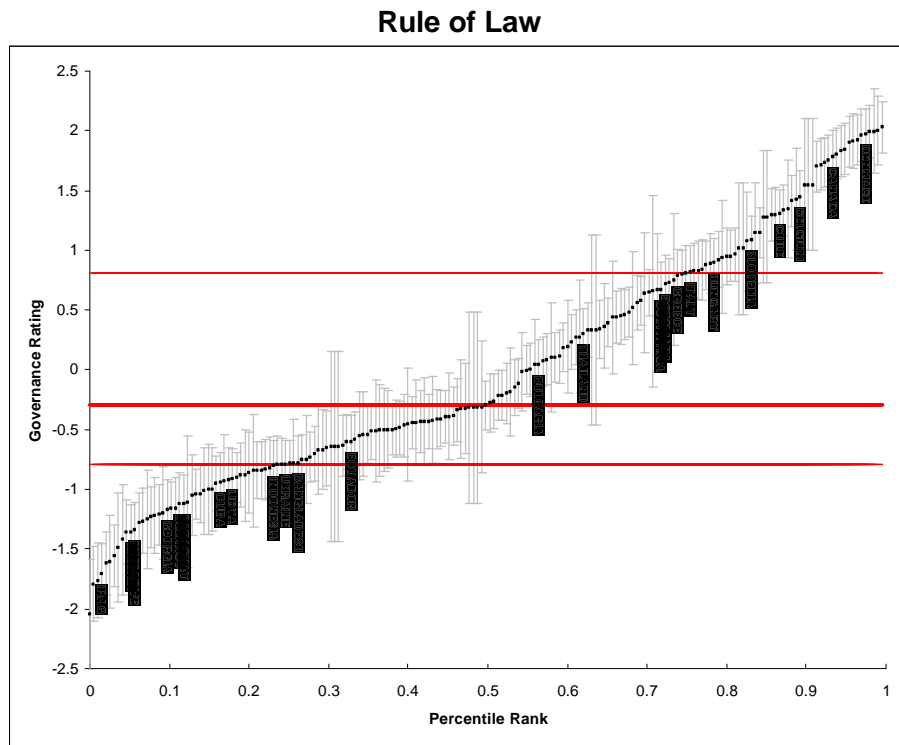
In conclusion, in the recent paper we have presented a substantially expanded and updated indicators of six dimensions of governance, based on several hundred individual measures of governance drawn from scores of sources and different organizations, covering the period 1996-2002. While we utilize the most efficient model for aggregation, resulting in lower margins of error than any alternative, we have emphasized throughout that such margins of error still remain substantial. These margins of error need to be taken seriously when comparing countries with each other and over time. This is especially the case when attempting to classify countries into groups according to their levels of governance for aid allocation purposes, as for example with the MCA.

We have addressed a number of important methodological issues relating to the construction and use of these governance indicators. We have argued that, for the purposes of measuring governance, there are few alternatives to the subjective, experiential data on which we rely. Moreover, in cases where objective indicators of governance are available, we have noted that these too have implicit margins of error,

and we have provided indicative calculations indicating that these margins of error are on the same order of magnitude as those associated with our subjective aggregates. We have also empirically investigated, and for the most part discounted, the importance of ideological biases in the perceptions data from polls of experts on which we rely. Finally, while our aggregate indicators measure countries' relative performance in each period, we have also examined the limited available evidence on trends over time in governance worldwide. Interpreting these trends is difficult, but we can state with some confidence that there is little if any evidence of improvements in global governance over the period we consider.

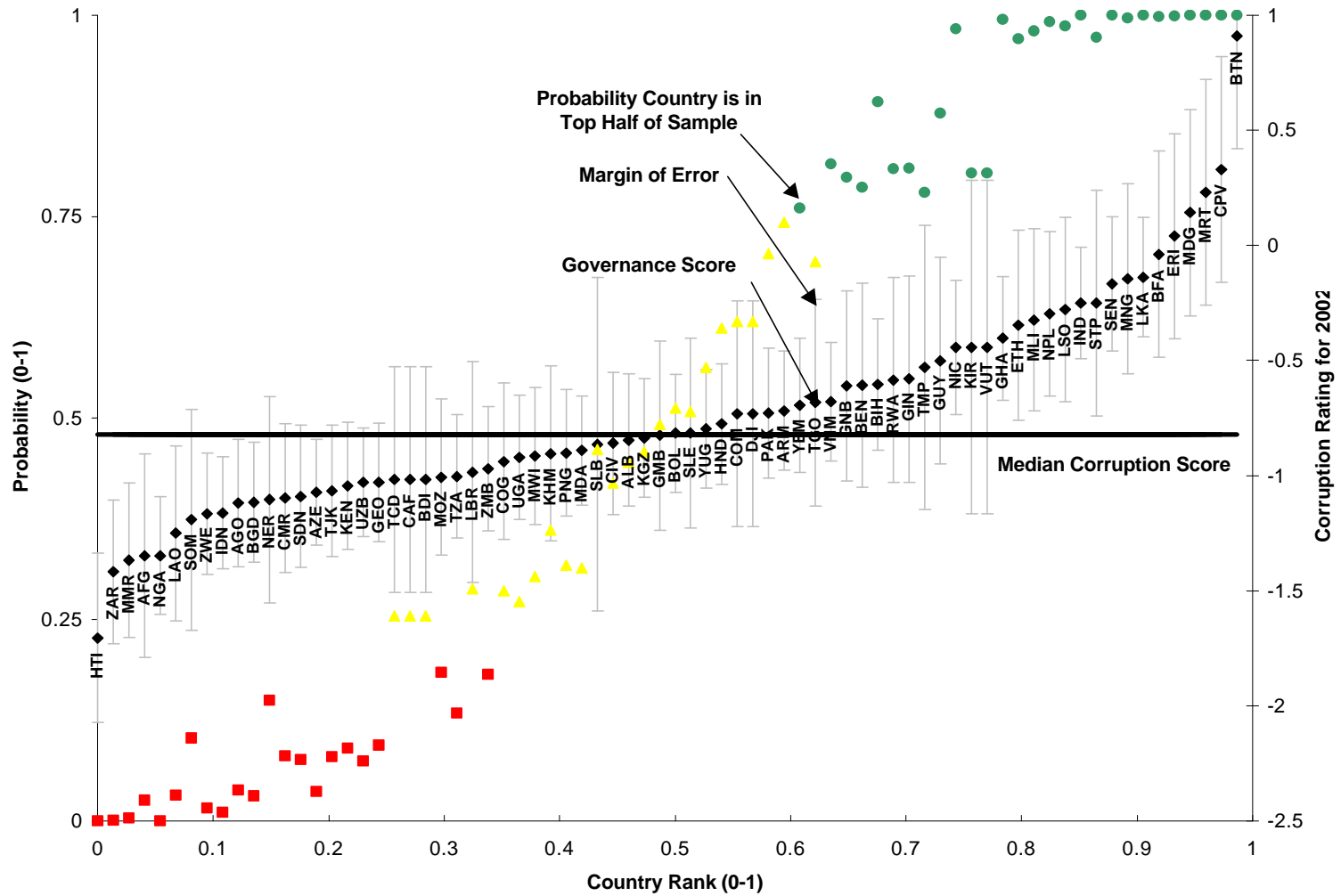
In continuing with this research project on measuring cross-country differences in governance, we expect that in the future the availability of additional data will enable further improvements in precision. The broader objective of this research project is to provide individual countries with a set of monitorable indicators of governance they can use to benchmark themselves against other countries and over time. We recognize however that there are limitations to what can be achieved with this kind of cross-country, highly-aggregated data. This type of data cannot substitute for in-depth, country-specific governance diagnostics as a basis for policy advice to improve governance in a particular country. Thus, a significant complementary effort has taken place to develop country-based [governance diagnostic](#) methodologies, based on in-depth surveys of enterprises, users of public services, and public officials. Such diagnostics permit the unbundling of governance performance by type of governance, by location, and by institution, permitting the use of the significant variation within a country to learn about the likely priorities for action for a country.

Figure 1: Estimates of Governance, 2002



Note: This graph shows estimates of the indicated dimension of governance (on the vertical axis) for all countries graphed against each country's percentile rank (on the horizontal axis) for 2002. The vertical bars show the statistically-likely range of values of governance for each country, with the midpoint of each bar corresponding to the best single estimate. Selected countries are labeled. As emphasized in the text, the ranking of countries along the horizontal axis is subject to significant margins of error, and this ordering in no way reflects the official view of the World Bank, its Executive Directors, or the countries they represent.

Figure 2: Using Governance Indicators to Allocate Aid



Note: This graph plots estimates of Control of Corruption in 2002 for all 74 countries potentially eligible for the first round of the MCA. Countries are ranked according to their corruption rating on the horizontal axis, and corruption scores are shown on the vertical axis. The vertical lines for each country report the 90% confidence intervals for corruption, and the midpoint of each interval indicates the corruption score. The red squares (yellow triangles) (green circles) report the probability that each country has a corruption score above the median. The colours correspond to three groups of countries for which this probability is less than 25%, between 25% and 75%, and above 75%. As emphasized in the text, the ranking of countries along the horizontal axis is subject to significant margins of error, and this ordering in no way reflects the official view of the World Bank, its Executive Directors, or the countries they represent.