

Governance Matters III: Governance Indicators for 1996-2002

Daniel Kaufmann
Aart Kraay
Massimo Mastruzzi

The World Bank

First Draft: June 30, 2003
Revised Version: April 5, 2004

Abstract. This paper presents 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 categories capturing key dimensions of governance, and use an unobserved components model to construct six aggregate governance indicators in each of the four periods. We present the point estimates of the dimensions of governance as well as the margins of errors for each country for the four periods. We also address various methodological issues, including tests for potential biases, and the interpretation and use of the data given the estimated margins of errors for our indicators as well as for others. The data, as well as a web-based graphical interface, are available at www.worldbank.org/wbi/governance/govdata2002/.

1818 H Street NW, Washington, D.C. 20433. dkaufmann@worldbank.org, akraay@worldbank.org, mmastruzzi@worldbank.org. The opinions expressed here, as well as the governance research indicators themselves, do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent. We would like to thank Marc Levy, Adrian Shute, Merli Baroudi, Guy Dunn, Jennifer Windsor, Rich Fullenbaum, Marta Lagos, Phil Keefer, Joel Hellman, Peter Cornelius, Fiona Pava, Rui Coutinho, Satish Mannan, Craig Webster, and David Cieslikowsky for providing data answering our numerous questions, Steve Radelet, Clay Lowery, and Steve Knack for their comments, and Erin Hoffmann for assistance. The support and collaboration of the U.S. State Department, the Netherlands Government, and Italian Trust Funds is appreciated.

1. Introduction

This paper presents a set of 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 variables measuring perceptions of governance, drawn from 25 separate data sources constructed by 18 different organizations. We assign the individual measures of governance perceptions to six categories capturing key dimensions of governance, and use an unobserved components model to construct six aggregate governance indicators in each of the four periods. The governance indicators reported here are an update and expansion of our previous work (Kaufmann, Kraay and Zoido-Lobaton (KKZ) 1999a,b and 2002 and Kaufmann and Kraay 2002). The data, as well as a web-based graphical interface, are available at www.worldbank.org/wbi/governance/govdata2002.

In this paper we first describe the data we draw on for this round of the governance indicators. As discussed in more detail below, we have obtained several new data sources, and we have updated our existing data sources forward and backward in time to 2002 and 1996 respectively. The indicators for 1998 and 2000 reported in this paper are based on the indicators for 1997/98 and 2000/01 that we constructed in previous work. However, in light of the availability of several new data sources, we revised these indicators to reflect the new data, resulting in the revised indicators for 1998 and 2000.

We next briefly review the aggregation procedure and describe the resulting governance indicators. 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 indicators relative to previous years, the margins of error associated with estimates of governance remain large relative to the units in which governance is measured. This implies 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 levels of governance.

We emphasize however that margins of error are not unique to the perceptions data we use: measurement error is pervasive among all measures governance and insitutional quality. Rather, we note that an advantage of our measures is that we are able to be explicit about the accompanying margins of error.

In the final section of the paper we discuss in more detail several issues related to the construction and use of our governance indicators. We first discuss the usefulness of subjective measures of governance relative to alternatives. We also empirically investigate the importance of ideological biases in expert assessments of corruption, and find little evidence that they are present. We then discuss in more detail the consequences of the substantial margins of error associated with the governance indicators, using the aid allocation rules proposed for the U.S. Government's Millennium Challenge Account as an illustration. We also note that margins of error are not unique to subjective indicators of governance, and show how to infer corresponding margins of error for more quantitative or objective indicators that have been proposed in recent years. We also provide some description of the limited available information on global trends in governance over the past six years covered by our indicators. Finally, we conclude with a comparison of our Control of Corruption indicator with the widely-used Corruption Perceptions Index produced by Transparency International.

2. Measuring Governance

2.1 Governance Clusters

We construct six aggregate governance indicators, motivated by a broad definition of governance as the traditions and institutions by which authority in a country is exercised. This includes (1) the process by which governments are selected, monitored and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions among them. This classification of indicators into clusters corresponding to this definition of governance is not intended to be definitive. Rather, drawing on existing definitions of governance, we unbundled it reflecting our views of what constitutes a consistent and useful organization of the data that is concordant with prevailing notions of governance.

The first two governance clusters are intended to capture the first part of our definition of governance: the process by which those in authority are selected and replaced. We refer to the first of these as “Voice and Accountability”, and include in it a number of indicators measuring various aspects of the political process, civil liberties and political rights. These indicators measure the extent to which citizens of a country are able to participate in the selection of governments. We also include in this category indicators measuring the independence of the media, which serves an important role in holding monitoring those in authority and holding them accountable for their actions. The second governance cluster is labeled “Political Stability and Absence of Violence”. In this index we combine several indicators which measure perceptions of the likelihood that the government in power will be destabilized or overthrown by possibly unconstitutional and/or violent means, including domestic violence and terrorism. This index captures the idea that the quality of governance in a country is compromised by the likelihood of wrenching changes in government, which not only has a direct effect on the continuity of policies, but also at a deeper level undermines the ability of all citizens to peacefully select and replace those in power.¹

The next two clusters summarize various indicators of the ability of the government to formulate and implement sound policies. In “Government Effectiveness” we combine responses on the quality of public service provision, the quality of the bureaucracy, the competence of civil servants, the independence of the civil service from political pressures, and the credibility of the government’s commitment to policies. The main focus of this index is on “inputs” required for the government to be able to produce and implement good policies and deliver public goods. The second cluster, which we refer to as “Regulatory Quality”, is more focused on the policies themselves. It includes measures of the incidence of market-unfriendly policies such as price controls or inadequate bank supervision, as well as perceptions of the burdens imposed by excessive regulation in areas such as foreign trade and business development.

¹ It is worth noting that there is some ambiguity regarding the normative direction of a few of the subcomponents this indicator. For example, a few of our sources rank countries such as Cuba and North Korea highly in terms of their political stability, which simply reflects the longevity of the governments in power in these countries.

The last two clusters summarize in broad terms the respect of citizens and the state for the institutions which govern their interactions. In “Rule of Law” we include several indicators which measure the extent to which agents have confidence in and abide by the rules of society. These include perceptions of the incidence of crime, the effectiveness and predictability of the judiciary, and the enforceability of contracts. Together, these indicators measure the success of a society in developing an environment in which fair and predictable rules form the basis for economic and social interactions, and importantly, the extent to which property rights are protected. The final cluster, which we refer to as Control of Corruption, measures perceptions of corruption, conventionally defined as the exercise of public power for private gain. Despite this straightforward focus, the particular aspect of corruption measured by the various sources differs somewhat, ranging from the frequency of “additional payments to get things done,” to the effects of corruption on the business environment, to measuring “grand corruption” in the political arena or in the tendency of elite forms to engage in “state capture”. The presence of corruption is often a manifestation of a lack of respect of both the corrupter (typically a private citizen or firm) and the corrupted (typically a public official or politician) for the rules which govern their interactions, and hence represents a failure of governance according to our definition.

2.2 Sources of Governance Data

In this section we provide a brief overview of the sources of governance data underlying our composite indicators. Table 1 provides a list of all of the sources that we have used to construct our governance indicators for 1996, 1998, 2000, and 2002.² Details on each of these sources, and on how we have assigned individual questions from each of these sources to our six governance clusters, are available in a data appendix at www.worldbank.org/wbi/governance/govdata2002/.

For the 2002 indicators we use 250 individual measures, taken from 25 different sources, produced by 18 different organizations. These organizations include international organizations, political and business risk-rating agencies, think-tanks, and

² Note we refer to the previously available periods, namely, 1997/98, as 1998, and 2000/01, as 2000, for convenience. As discussed in this paper, the estimates for these two previously available periods have also been updated and revised, given the additional data that became available.

non-governmental organizations. In 2002, we have included six new data sources relative to our previous work: Afrobarometer (a survey of individuals in 12 African countries), Reporters Without Borders (an assessment of press freedoms compiled by an international journalists' organization), Human Rights (a numerical coding of the assessments of certain dimensions of human rights as reported by U.S. State Department and Amnesty International, as first reported in Cingranelli and Richards (2001) and subsequently updated and expanded by Craig Webster, World Markets Online (a commercial risk-rating agency), Voice of the People (a citizen survey sponsored by Gallup International), and the World Bank's Country Policy and Institutional Assessment (known as CPIA), an internal assessment of country performance constructed by World Bank country economists.³

Two of the new sources that we have added in 2002 are also available for earlier years (Human Rights and CPIA). In order to make full use of this additional data, as well as to improve the comparability of the governance indicators over time, we have revised our previous indicators for 1997/98 and 2000/01 to incorporate these sources. In particular, we begin with the same underlying sources that we used in the past for these two periods. We then add the data from Human Rights and CPIA, and we discard two minor sources that we have used in the past but are no longer available.⁴ We then recompute the aggregate indicators based on this revised dataset. For notational convenience, we refer to these revised indicators as the indicators for 1998 and 2000, even though both of these measures are based on data from a two-year period. Finally, a subset of our indicators are available in 1996. Based on these sources, we also construct aggregate governance indicators for this period.

As in our previous work, we continue to distinguish sources according to (1) whether they are *polls* of experts, or *surveys* of businesspeople or citizens in general,

³ It is worth noting that we do not use the Transparency International Corruption Perceptions Index (CPI) as a component of our aggregate corruption indicator. This is because the CPI is itself an aggregate of a number of individual sources, all of which we have already included in our corruption indicator.

⁴ These are the Central European Economic Review which rated a sample of transition economies and ceased publication after our only use of this source in the 1997/98 indicators, and the Political and Economic Risk Consultancy, which has also discontinued its rating of a small number of Asian economies. Dropping these sources does not affect our country coverage, and makes the aggregate indicators more comparable over time.

and (2) the extent to which the sample of countries included in the sources is *representative* of the world as a whole. The advantages and disadvantages of polls of experts relative to surveys of market participants are well-known, and are discussed in our previous papers (KKZ 1999b, 2002 and Kaufmann and Kraay 2002). Briefly, the choice among these two types of governance data involves tradeoffs in terms of cross-country comparability versus first-hand knowledge of local conditions. Polls of experts in general are explicitly designed to provide comparable results across countries, through elaborate benchmarking procedures. However, their reliability depends greatly on the ability of the small group of experts involved to provide objective and accurate assessments of the governance dimensions being measured.⁵ In contrast, surveys typically draw on the responses of large numbers of local respondents with direct first-hand knowledge of local conditions. However, to the extent that ostensibly identical survey questions are interpreted differently by respondents with different cultural and/or socioeconomic backgrounds, it can be difficult to make cross-country comparisons using survey responses.⁶ For example, it is easy to see how responses to a question on the prevalence of “improper practices” can be affected by country-specific perceptions of what such practices might be.

Sources of governance data also vary with respect to the sample of countries they cover. A number of sources cover a very large sample of developed and developing countries, while others cover very narrowly-focused samples of countries. Also, many of the poorest and/or smallest countries in the world tend not to be covered by many commercially-oriented polls because they are relatively unattractive to foreign investors.

⁵ Most of the polls of experts we rely on are based on the views of a large group of raters. For example, the Economist Intelligence Unit is based in London and prepares its assessments based on the views of a worldwide network of correspondents. Similarly, Freedom House is based in New York but bases its assessments on a global network of correspondents and human rights monitors, and Reporters Without Borders, based in France, constructs its assessments based on a global network of reporters. Other polls of experts have a narrower institutional affiliation for their respondents. For example the European Bank for Reconstruction and Development (EBRD) Transition Report ratings are primarily based on the assessments of EBRD staff based in London and the State Department component of our Human Rights measure reflects the views of U.S. State Department employees. Below we discuss in more detail evidence regarding potential ideological biases in the assessments of polls of experts.

⁶ It is important to note that our three main sources of firm-level survey data (the Geneva-based World Economic Forum’s Global Competitiveness Report, the Lausanne-based Institute for Management Development’s World Competitiveness Yearbook, and the Washington-based World Bank’s business environment surveys) primarily interview domestic as opposed to foreign-owned firms in the countries they cover.

Since there is a strong positive association across countries between governance and per capita incomes, this difference between sources makes it difficult a priori to compare indicators from sources which cover sets of countries with very different income levels. Similarly, there may be regional differences in governance which hamper simple comparisons across sources. For example, it is difficult to compare a governance rating based only on transition economies with one based on a broad set of countries. As discussed in a previous paper (KKZ 1999a), the methodology we use to construct aggregate governance indicators takes these differences in country coverage into account as the data from individual sources are transformed into common units for aggregation across sources. Table 1 also contains a column indicating which sources we treat as representative and non-representative.

2.3 Aggregation Methodology

Implicit in our organization of the data is the view that, within each cluster, each of these indicators measures a similar underlying basic concept of governance. Given this view, there are considerable benefits from combining these related indicators into an aggregate governance indicator for each cluster. First, the aggregate indicators span a much larger set of countries than any individual source, permitting comparisons of governance across a broader set of countries than would be possible using any single source. Second, aggregate indicators can provide more precise measures of governance than individual indicators. Third, it is possible to construct quantitative measures of the precision (and thus margins of error) of both the aggregate governance estimates for each country, as well as their components.

We now present in some detail the statistical methodology used to construct our aggregate governance indicators.⁷ For each of the governance clusters, we combine the component indicators into an aggregate governance indicator using the same methodology used to calculate our first set of indicators, as documented in detail in KKZ (1999a). We use an extension of the standard unobserved components model which expresses the observed data in each cluster as a linear function of the unobserved common component of governance, plus a disturbance term capturing perception errors

⁷ Readers who are already familiar with our approach, or who are not interested in the technical details, can skip the remainder of this subsection.

and/or sampling variation in each indicator.⁸ In particular, we assume that we can write the observed score of country j on indicator k , $y(j,k)$, as a linear function of unobserved governance, $g(j)$, and a disturbance term, $\varepsilon(j,k)$, as follows:

$$(2.1) \quad y(j,k) = \alpha(k) + \beta(k) \cdot (g(j) + \varepsilon(j,k))$$

where $\alpha(k)$ and $\beta(k)$ are unknown parameters which map unobserved governance $g(j)$ into the observed data $y(j,k)$. As a choice of units, we assume that $g(j)$ is a random variable with mean zero and variance one. We assume that the error term has zero mean and a variance is the same across countries, but differs across indicators, i.e. $E[\varepsilon(j,k)^2] = \sigma_\varepsilon^2(k)$. Finally we assume that the errors are independent across sources, i.e. $E[\varepsilon(j,k) \cdot \varepsilon(j,l)] = 0$ for l different from k . This assumption imposes the identifying assumption that the only reason why two sources might be correlated with each other is because they are both measuring the same underlying unobserved governance dimension.⁹

The disturbance term $\varepsilon(j,k)$ captures two sources of uncertainty in the relationship between true governance and the observed indicators. First, the particular aspect of governance covered by indicator k is imperfectly measured in each country, reflecting either perception errors on the part of experts (in the case of polls of experts), or sampling variation (in the case of surveys of citizens or entrepreneurs). Second, the relationship between the particular concept measured by indicator k and the corresponding broader aspect of governance may be imperfect. For example, even if the particular aspect of corruption covered by some indicator k , (such as the prevalence of “improper practices”) is perfectly measured, it may nevertheless be a noisy indicator of corruption if there are differences across countries in what “improper practices” are

⁸ Unobserved components models were pioneered in economics by Goldberger (1972), and the closely-related hierarchical and empirical Bayes models in statistics by Efron and Morris (1971, 1972).

⁹ For some pairs of sources, this assumption may not be literally true. For example, it will be violated if different risk rating agencies base their own assessments on the assessments of other agencies included in our sample. We have to the best of our knowledge excluded any source of governance data where we found that it was explicitly based on another one of our sources. Nevertheless, the possibility of correlated errors remains. The main consequence of this is that our standard errors will be biased downwards -- see Kaufmann, Kraay and Zoido-Lobaton (1999a) for an example. This underscores the importance of caution in comparing governance estimates across countries and over time, that we emphasize throughout.

considered to be. Both of these sources of uncertainty are reflected in the indicator-specific variance of the error term, $\sigma_{\varepsilon}^2(k)$.

Given estimates of the parameters of the model, $\alpha(k)$, $\beta(k)$, and $\sigma(k)$, we can compute estimates of governance for each country, as well as measures of the precision of these estimates. Formally, the estimate of governance for a country produced by the unobserved components model is the mean of the distribution of unobserved governance conditional on the $K(j)$ observed data points for that country. This conditional mean is the following weighted average of appropriately-rescaled scores of each of the component indicators:

$$(2.2) \quad E[g(j) | y(j,1), \dots, y(j, K(j))] = \sum_{k=1}^{K(j)} w(k) \cdot \frac{y(j,k) - \alpha(k)}{\beta(k)}$$

where the weights applied to each source k , $w(k) = \frac{\sigma_{\varepsilon}(k)^{-2}}{1 + \sum_{k=1}^{K(j)} \sigma_{\varepsilon}(k)^{-2}}$, are inversely

proportional to the variance of the error term of that source. As we discuss in more detail in the final section of the paper, we find that the efficiency gains from precision-weighting are substantial relative to the alternative of simply averaging re-scaled scores from each source for each country. We also report the standard deviation of this conditional distribution as an indicator of the confidence we can have in this estimate, which is:

$$(2.3) \quad SD[g(j) | y(j,1), \dots, y(j, K(j))] = \left(1 + \sum_{k=1}^{K(j)} \sigma_{\varepsilon}(k)^{-2} \right)^{-\frac{1}{2}}.$$

This standard deviation is declining in the number of individual indicators in which a particular country appears, and is increasing in the variance of the disturbance term on each of these indicators.

The assumptions of the unobserved components model ensure that the distribution of unobserved governance in each country is normal, conditional on the data for that country. Therefore, these conditional means and standard deviations for each country have a natural interpretation. For example, a useful interpretation of the reported

estimates and standard deviations for each country is to note that there is a 90% probability that the “true” level of governance in a country is in an interval of plus or minus 1.64 times the reported standard deviation centered on the point estimate itself. We refer to such a range as a 90% confidence interval around the estimate of governance for a country.¹⁰

In order to implement this approach, we require estimates of all of the unknown survey-specific parameters, $\alpha(k)$, $\beta(k)$, and $\sigma_{\varepsilon}^2(k)$. We do this in a two-stage procedure. First, we assume that governance and the error terms in Equation (2.1) are jointly normally distributed, and then apply maximum likelihood methods using only the representative sources to retrieve the parameters of interest for each governance cluster. This is nothing more than a standard application of the unobserved components model. We cannot however include our many non-representative sources in the first stage of the estimation procedure. This is because the distribution of unobserved governance in the subset of countries covered by these surveys is not the same as that in the world as a whole. As a result, for these sources we cannot make the assumption that unobserved governance in the countries covered by these surveys follows a standard normal distribution, as is required by the maximum likelihood procedure.

We instead obtain the parameters of the non-representative sources in a two-stage procedure. First, we estimate the parameters of the model only for the representative sources, and construct preliminary estimates of governance based only on these sources. In the second stage, we treat these preliminary estimates as an observable proxy for governance, and obtain the parameters of interest for the non-representative sources by regressing these indicators on observable governance, i.e. by directly estimating Equation (2.1).¹¹ We then use all the estimated parameters of the unobserved components model to construct a final set of estimates of governance.

¹⁰ This is a slight abuse of terminology, as these are not confidence intervals in the usual frequentist sense of a stochastically varying interval centered around a fixed unknown parameter. Rather, we treat governance as a random variable, and the 90% confidence interval is simply the 5th and 95th percentiles of the conditional distribution of governance given the observed data.

¹¹ In order to get consistent estimates of the parameters of the non-representative sources, we need to adjust for attenuation bias caused by the fact that our observable proxy for governance is a noisy indicator of true governance. Fortunately, we can use the information on the standard errors associated with the governance estimates obtained in the first stage to do this.

The resulting estimates of governance have an expected value (across countries) of zero, and a standard deviation (across countries) of one. Due to sampling variability, this will not be exactly true for any one of our governance indicators in any period. To avoid any confusion regarding the units of the governance indicators, we rescale the estimates of governance by subtracting the mean (across countries) and dividing by the standard deviation (across countries) for each indicator, so that each indicator has a mean of zero and a standard deviation of one in each period.

It is also important to note that we have assumed that the distribution of unobserved governance is the same in every period. In particular, this imposes the restriction that the mean or world average of governance is the same in each period. As a result, our indicators are not informative about global trends in governance, although they are potentially informative about changes in countries' relative positions over time. In the final section of this paper we discuss the limited available information on global trends in governance, based on some of our underlying sources.

3. Governance Indicators 1996-2002

3.1 Levels of Governance Worldwide

The aggregate governance indicators resulting from our methodology, for all countries, for each of the six indicators and for all four periods, are reported in Appendix C, Tables C1-C6. The governance estimates are normally distributed with a mean of zero and a standard deviation of one in each period. This implies that virtually all scores lie between -2.5 and 2.5, with higher scores corresponding to better outcomes. As discussed in the previous section, this also implies that our aggregate estimates convey no information about global trends in governance.¹² They are however informative about changes in countries' relative positions over time. Later in the paper we will discuss the limited available evidence on trends over time in global averages of governance.

Table 2 summarizes some of the key features of our governance indicators. In the top panel we show the number of countries included in each of the six indicators and

¹² However, analyzing some of the individual sources can provide insights on such global trends, as we discuss in Section 4.5 below.

four periods. The Voice and Accountability indicator covers the largest set of 199 countries.¹³ With the exception of Political Stability which covers 186 countries, all the other indicators cover 195 countries. Over time, there has been a steady increase in the number of countries covered in each of our indicators. This increase in the number of data sources is reflected in an increase in the median number of sources available per country, which, depending on the governance component, ranges from four to six in 1996, and from six to eight in 2002. The proportion of countries in our sample for which our governance estimates are based on only one source has also declined considerably, to an average of only 10 percent of the sample in 2002. In the 2002 indicators we now cover virtually all countries in the world, and thus we do not expect major future improvements in country coverage.

An important consequence of this expanded data availability is that the margins of error for the governance indicators have declined, as shown in the final panel of Table 2. Depending on the governance component, in 1996 the average (for all countries) of the margins of error ranged from 0.26 to 0.40, while in 2002 the corresponding range is from 0.19 to 0.27. Moreover, the average margins of error for the revised 1998 and 2000 indicators are also lower than the previously available estimates, again reflecting the incorporation of more data for more countries. These declines in margins of error illustrate the benefits in terms of precision of constructing composite indicators based on as much information as possible.

Despite this increase in precision and the benefits of aggregation, the margins of error associated with estimates of governance remain substantial relative to the units in which governance is measured. We illustrate this point in Figure 1. In each panel of Figure 1, we order countries in ascending order according to their point estimates of governance in 2002 on the horizontal axis, and on the vertical axis we plot the estimate of governance and the associated 90% confidence interval described above. The size of these confidence intervals varies across countries, as different countries appear in different numbers of sources with different levels of precision. The resulting confidence intervals are large relative to the units in which governance is measured. To emphasize

¹³ A few of the entities covered by our indicators are not independent states (Puerto Rico, Hong Kong, West Bank/Gaza, Martinique, and French Guyana). A handful of very small independent principalities (Monaco, San Marino, and Andorra) are also included. For stylistic convenience we will refer to all 199 entities as “countries”.

this point, the horizontal lines in Figure 1 delineate the quartiles of the distribution of governance estimates. Even though the differences between countries in the bottom and top quartiles are substantial, the number of countries that have 90% confidence intervals that lie entirely within a given quartile is not large.

From Figure 1 it should also be evident that many of the small differences in estimates of governance across countries are not likely to be statistically significant. For many applications, instead of merely observing the point estimates, it is therefore more useful to focus on the *range* of possible governance values for each country (as summarized in the 90% confidence intervals shown in Figure 1). For two countries at opposite ends of the scale of governance, whose 90% confidence intervals do not overlap, it is clear that there are in fact significant differences in governance between these two countries. For pairs of countries that are closer together and whose 90% confidence intervals overlap (which can also be seen in Figure 1), one should be much more circumspect about the significance of estimated differences in governance between two such countries.

Finally we consider the reasons for the differences across countries in the margins of error associated with governance estimates. As noted above, these differences are due to two factors: (i) cross-country differences in the number of sources in which a country appears, and (ii) differences in the precision of the sources in which each country appears. These two sources of variation can be seen from Equation (2.3), which shows that the precision of our estimates of governance increases as the number of sources increases, and as each underlying source becomes more precise. Across countries, the standard error of the governance estimate for a country declines at the rate of the inverse of the square root of the number of sources, consistent with our assumption that errors are uncorrelated across sources. What this means in practice is that a 90% confidence interval for a country with only one source will be roughly twice as large as the 90% confidence interval for a country appearing in the median number of seven sources.

The second reason why margins of error differ across countries is because our sources differ in their estimated precision, and different countries appear in different sets of sources. Recall that for each source, we estimate the variance of the error with which

it measures the unobserved “true” level of governance across countries. We then use the inverse of these estimated variances to weight sources when constructing the aggregate score for each country. This means that more precise sources (in the sense of providing less noisy signals of governance) receive more weight in the aggregate indicators. This procedure minimizes the variance of our estimates of governance for each country. Table 3 summarizes, for all sources, indicators, and periods, the weights applied to each source in constructing the corresponding aggregate indicator. In particular, the entries in each column report the weights that would be assigned to each source, for a hypothetical country appearing in all sources. If a country appears in a subset of sources, the weights applied would be proportionate to the ones reported for those sources. Looking across sources, we see that there is a great deal of variation in the weights assigned to different sources, reflecting substantial differences in our estimated precision of each individual source.¹⁴ These differences are in turn reflected in differences in margins of error associated with governance scores for each country.

3.2 Changes over Time in Estimates of Governance

We now briefly discuss the changes over time in the governance indicators we have constructed. Before discussing specific countries and indicators, it is useful to distinguish between several potential reasons for changes in observed estimates of governance over time. Mechanically, the observed change in governance for a given country between two points in time can be attributed to four distinct factors: (1) changes in the perceptions of governance recorded in the underlying sources available in both periods for that country; (2) changes in the weights the aggregation procedure applies to different sources in each period; (3) additions to or deletions from the set of sources in which a country appears; and (4) the addition of new countries to the aggregate indicator which systematically rate better or worse than the country in question (since as discussed above our indicators measure only countries’ relative positions).

¹⁴ When we estimate the unobserved components model, the estimated precision of each source reflects the extent to which that source is correlated with other sources. In our empirical framework, we are assuming that errors are uncorrelated across sources. As a result, sources that tend to be highly correlated with other sources are more informative, and hence have lower error variances, than sources that tend to be only weakly correlated with other sources.

Changes in scores assigned to countries by underlying sources is the most direct and straightforward measure of how governance in a country changes over time. Interpreting changes over time in the aggregate indicators due to the remaining three reasons is somewhat more difficult, and involves tradeoffs. Consider first the re-weighting of sources that occurs from year to year as the observed correlations among sources change. The advantage of this re-weighting is that it ensures that in each year we have the most precise possible estimates of the level of governance. On the other hand, the disadvantage is that some of the changes over time in governance estimates for a given country will reflect this re-weighting rather than changes in the underlying indicators. However, in a previous paper (KKZ 2002) we showed that these changes in weights on average account for only a small fraction of the variance of changes in governance estimates.

There are also tradeoffs in interpreting the changes in governance estimates due to the addition of new sources for that country. Adding new data sources improves the precision of our estimates of the *level* of governance in a country at a single point in time. However, if the new source or sources rate a country significantly differently from existing sources, this can result in changes in estimates of governance which need not reflect actual changes in governance, but rather simply reflect the inclusion of new information relative to the previous period. In order to reduce the importance of this source of variation in governance estimates, as well as to improve the precision of our estimates of the level of governance in past years, we have recalculated our previous indicators incorporating all of the data we now have at our disposal. Nevertheless, the 2002 indicators also reflect the information embodied in a few new sources relative to 2000 and previous periods, and this provides a further reason why changes over time should be interpreted with some caution.¹⁵

Since our indicators measure only countries' governance performance relative to each other, changes over time in relative performance may also reflect the addition of new countries to the aggregate indicator. If for example we add a country with a governance rating that is high relative to those countries already in the index, then by construction all the countries which rank lower than this country will receive lower

¹⁵ On the website displaying the data (www.worldbank.org/wbi/governance/govdata2002/) we have made available an option for users to identify the sources of governance data used for each country, indicator, and period.

scores. It is easy however to remove this source of bias from comparisons of governance estimates over time. For example, if comparisons over time are based on changes in countries' percentile ranks within the set of countries common to both periods, then this source of variation is removed. We therefore recommend this procedure as a robustness check when considering changes over time in a specific country or set of countries. In practice, however, we find that this source of bias is relatively small, especially when comparing 2002 with 2000, since following the revisions for the 2000 estimates due to the increased availability of sources there are now only small changes in the number of countries covered between these two periods.

A final issue concerns the statistical significance of observed changes in the aggregate indicators. The basic observation is that *changes* in the estimates of governance tend to be small relative to the *levels* of governance. It is however difficult to be more precise about the statistical significance of changes in governance, for the following reason. Recall that for each period, our aggregation procedure allows us to summarize our knowledge about governance in a given country in terms of the distribution of unobserved governance conditional on the data that we see for that country. As discussed above, we use the mean of this conditional distribution as our best estimate of the level of governance in a country, and the standard deviation of this distribution to summarize the precision of our knowledge about governance for that country. However, when we repeat our aggregation procedure in successive periods, we do not produce any information about the *joint* distribution of governance in successive periods. Without this joint distribution, we cannot make precise probabilistic statements about changes over time in governance.¹⁶

Instead, we adopt the simple rule of thumb of focusing only on changes in governance for those countries in which the 90% confidence intervals in the two periods do not overlap. Figure 2 illustrates the use of this rule of thumb to highlight large changes in governance between 2002 and 2000. In each of the six panels, we plot the 2002 score on the horizontal axis, and the 2000 score on the vertical axis. We also plot the 45-degree line in red, so that countries above this line correspond to declines in the quality of governance, while countries below the line correspond to improvements in

¹⁶ Extensions of our aggregation procedure along the lines of dynamic unobserved component models could in principle provide information about the joint distribution of governance over time. We have not yet attempted to implement this idea with our data.

governance. The countries with changes in governance that are large relative to their margins of error in each period are highlighted in red, and the 90% confidence intervals in both periods are indicated by vertical and horizontal lines. Finally, for each country which appears in the 2002 indicator but not in the 2000 indicator, we plot its score along the 45-degree line. This gives a visual summary of the distribution of governance among those countries that we have added to the sample in 2002.

The first striking feature to note from Figure 2 is that the number of countries with large changes in governance over this brief period is quite small, ranging from just one in the case of Voice and Accountability, to a maximum of 11 in the case of Regulatory Quality. This should not be surprising given the relatively short time period under consideration and the reality that in most countries, changes in governance tend to be quite gradual. Many of these changes are understandable in light of developments over the past two years. For example Argentina's recent financial crisis is reflected in strong declines in perceptions of governance across the board. Similarly, the recent turmoil in Zimbabwe is associated with a sharp decline in perceptions of Rule of Law. An interesting case is the United States, which registers a decline in Political Stability and Absence of Violence, reflecting heightened concerns about terrorism in the aftermath of September 11, 2001. On the other hand, reductions in sectarian violence in Sri Lanka drive an improved score in this category.

The reasons for changes in some of the other countries highlighted in Figure 2 are less obvious. We examine these in more detail in Table 4. The first two columns of this table report the governance scores in 2000 and 2002. The next four columns summarize the importance of changes in a country's scores on the underlying indicators in which it appears. In the column labeled "Agree" we report the number of sources available in both periods which move in the same direction as the aggregate indicator. The columns labeled "No Change" and "Disagree" report the number of sources on which that country's score does not change or moves in the opposite direction to the aggregate indicator. For each country we also summarize the extent to which changes in the individual sources agree with the direction of change in the aggregate indicator by calculating the "Agreement Ratio", or "Agree" / ("Agree" + "Disagree"). Finally, the last two columns summarize the changes in the set of sources on which each country's

score is based, reporting the number of sources dropped and added between the two periods.

Table 4 shows that the agreement ratio is quite high for countries with large changes in governance. Averaging across all countries and indicators, we find an average agreement ratio of 0.79. This provides some confidence that for countries with large changes in our governance estimates, these changes are being driven primarily by changes in underlying sources. In fact, there are only four countries where the agreement ratio is less than one-half – Belarus and Iraq for Regulatory Quality, and Madagascar and West Bank/Gaza for Control of Corruption. Noting that they are exceptions to the more generalized finding of broad agreement across sources, it is nonetheless useful to address transparently these anomalous cases. In the case of Belarus, the surprisingly high score in 2002 is driven primarily by its very strong responses from firms in BPS. Iraq on the other hand illustrates an unusual case in which the re-weighting of sources has a substantial effect on changes over time. In both periods, Iraq has one of the very worst scores in the world, so clearly the large change in Iraq's score does not reflect any real improvement during the period under consideration. Instead, it reflects the fact that the source which rated Iraq highest in 2002 received much less weight in the 2000 aggregate. For Control of Corruption, the large improvement observed in Madagascar and the large decline in West Bank/Gaza are both driven entirely by changes in the set of sources in which these countries appear. Madagascar appears in one new source which rates it quite highly (WMO), while it does not appear in the 2002 version of CDU which gave it a poor score in 2000. Conversely for West Bank/Gaza, it fared quite well on WBS in 2000, but this source is not available in 2002 and it scores poorly on the only source we have available for 2002, WMO.

In summary, Table 4 suggests that at least for large changes in governance, we can have relative confidence that these are primarily being driven by changes in these countries' scores on the indicators we have available for both periods. It is important to note however that the majority of smaller changes in governance reflect a combination of all four sources of variation discussed above. This can be seen from Table 5 which repeats the information in Table 4, but instead reports averages across all countries for the typically much smaller changes in the six governance indicators, between 2002 and

2000. Looking at all changes, rather than only at large changes as in the previous table, we see that the agreement ratio varies from 0.57 to 0.64, and for all six indicators is substantially lower than the agreement ratio for large changes in Table 4. This suggests that a greater degree of caution is in order when interpreting the typical small changes in governance estimates that we see from one period to the next. In principle we would expect changes over longer periods of time, such as 1996-2002, to be more informative. However, this is in part offset by the fact that there are larger changes in the composition of the sources comprising the aggregate indicators between these two periods.

4. Uses and Limitations of Governance Indicators

4.1 Perceptions Matter

In this subsection we briefly discuss why we rely exclusively on subjective or perceptions-based measures of governance when constructing our aggregate governance indicators. The primary reason for this choice is that for many of the key dimensions of governance, such as corruption or the confidence that property rights are protected, relevant objective data are almost by definition impossible to obtain, and so there are few alternatives to the subjective data on which we rely.

Consider for example corruption. Since corruption is by nature an illegal activity, direct measures of its prevalence do not exist. A variety of indirect measures are possible, but none are without difficulty. For example, relying on the frequency of references to corruption in the media will reflect not only the prevalence of corruption, but also the extent to which the press are free and objective in their coverage of events. Similarly, relying on prosecutions or conviction rates in corruption trials will to no small extent reflect the competence and independence of the police and judicial system, and thus will not exclusively reflect the prevalence of corruption itself. Finally, in recent years a handful of papers have attempted to measure corruption by looking for patterns in objective data that can only be consistent with corruption. For example, DiTella and Shargrodsky (2003) document variation in the procurement prices paid for very homogenous medical inputs such as syringes across hospitals in Buenos Aires as an indicator of corruption in procurement. Along similar lines, Golden and Picci (2003) carefully document variation in the differences between existing stocks of public

infrastructure and past flows of infrastructure spending across Italian regions, interpreting this gap as a measure of procurement corruption. While these last two papers represent important and interesting developments in measurement, cross-country measures of corruption based on this idea are not available – nor are they likely to be, given the major data requirements for this kind of exercise.¹⁷

For some other dimensions of governance, objective measures may be available, but nevertheless suffer from two related weaknesses. For Voice and Accountability, it is possible to use objective data on the presence of elections to measure democratic participation. However, there is a great deal of variation across countries in the extent to which the outcome of elections actually reflect the will of the voters. Measuring the extent to which elections are subverted, either through intimidation, manipulation, or sheer fabrication of results, brings us quickly back to the realm of more subjective or perceptions-based data. This is just one example of the important distinction between *de jure* and *de facto* situations regarding governance across countries. Countries may have extensive formal protections of property rights codified in their legal system, that are honored only in the breach. For example, most countries in the world now have formal independent anti-corruption commissions, but their effectiveness varies greatly.¹⁸

More generally, subjective perceptions of governance often matter as much as the legal reality. For example, on the basis of firm's perceptions on the undue influence on the political decision-making process exerted by powerful firms, influencing laws, policies and regulations, Hellman and Kaufmann (2003) develop a measure for 'crony bias' or unequal influence across firms. The authors find a consistent pattern in which perceived unequal influence has strongly negative impact on the firm's assessment of public institutions, which in turn affects the behavior of the firm towards those institutions. Crony bias at both the firm and the country level is associated with lesser use of the courts by the firms to resolve business disputes, with lower enforceability of court decisions, lower levels of tax compliance, and higher levels of bribery. Thus, the evidence suggests that the inequality of influence not only damages the credibility of

¹⁷ Furthermore, typically these within-country measures that are based on prices, assets and expenditure patterns are a proxy of the combined effect of the extent of mismanagement, inefficiency, and corruption. Disentangling the pure effect of the latter is far from simple.

¹⁸ Similarly, many countries have instituted anticorruption agencies, with very mixed results in terms of concretely addressing corruption.

institutions among less (politically) powerful firms, but affects the likelihood that they will use and provide tax resources to support such institutions, thereby perpetuating the weakness of such institutions and likelihood of capture by the influential.

Finally, in recent years the economics and comparative political economy literature has generated a profusion of results linking a variety of objective measures of the structure of institutions to a range of governance outcomes. A non-exhaustive list of examples includes the links between decentralization and corruption; the effects of the structure of the legal system on financial market development; the effect of checks and balances in the political system on regulatory and fiscal performance; the effects of democratic institutions on a wide range of socioeconomic outcomes; and many others. While this literature has served to greatly expand our understanding of the deep institutional determinants of development, the objective measures of institutional quality and/or their historical determinants on which they rely do not lend themselves well to the construction of aggregate governance indicators like ours. The basic reason is that these indicators typically do not have normative content on their own, but only do so in the context of a particular empirical analysis linking these variables with a particular outcome. For example, while measures of decentralization may be correlated with the incidence of corruption across countries, generally the explanatory power of this variable is not sufficiently strong that decentralization could be considered to be a reasonable proxy for corruption.

None of this is to suggest that the subjective data on which we rely are problem-free. We have already discussed the relative strengths and weaknesses of polls of experts and stakeholder surveys in measuring governance. Beyond this, a generic problem with many perceptions-based questions about governance is that they can be vague and open to interpretation. For example, a well crafted question to enterprises on corruption asks them for the estimated share of bribes in revenues expended annually by firms like theirs, and similarly another focused 'experiential' question probes into the percentage of the firm's management time spent dealing with government officials on red tape. By contrast, generalized opinion questions such as a citizen's perception of the overall tolerance of the population to corruption are less informative for our purposes. Nowadays we can increasingly rely on more specific, better crafted, and to an extent experiential questions, thanks to improvements that have taken place over time. For

instance, in contrast with the mid-nineties, the GCS survey of firms contains much more specific questions to the firm about corruption and governance, and some are of a quantitative and experiential nature (such as percentage of senior management time spent with public officials); similarly BPS includes many detailed questions unbundling governance to very specific components, and quantifying phenomena such as the percentage of bribes paid yearly as a share of the firm's revenues.

4.2 Potential Ideological Biases

A potential drawback of subjective data from polls of experts is that this kind of data may reflect the ideological tendencies of the institutions compiling the performance ratings. Our prior has been that this is not a major concern for the sources on which we rely. This is because we find a very high degree of correlation among virtually all of our sources, which is difficult to reconcile with a systematic ideological bias among certain sources. Nevertheless, it is useful to investigate the extent to which the differences in assessments across sources are related to observable measures of the ideology of the government in power in each country.

We do this as follows. Our identifying assumption is that surveys of firms or individuals are not tainted by ideology, since they reflect the views of a large number of respondents in each country. In contrast, it is possible that the views of a smaller number of raters affiliated with a particular institution may reflect the ideology of that group. We can therefore identify the effects of ideology by looking at the correlation across countries between the ideology of the government in power, and the *difference* in the percentile ranks assigned to countries by a poll of experts and a survey of individuals and firms. We implement this idea using the World Bank's Business Environment Survey (WBS) for 2000, and an independently available indicator variable that takes on the value 1 if the government in power is left-of-center, 2 if it is center, and 3 if it is right-of-center, taken from the database of political institutions constructed by Beck et. al. (2001).¹⁹ The coefficient on the ideology variable will therefore capture the extent to which a given poll of experts rates countries with left- or right-wing governments systematically differently from a survey.

¹⁹ Since most of the countries are coded as left or right, we obtain almost identical results if we instead include dummy variables for left- and right-wing governments.

The results are summarized in Table 6. The columns of Table 6 correspond to each of our polls of experts. In each of the six panels, we show the results of a regression of the difference between the percentile rank of a country on that poll and its rank on WBS on the indicator variable measuring the ideology of the government in that country. All variables are measured in 2000, which is the most recent year for which the ideology variable is available.²⁰ A positive coefficient indicates that the poll in question tends to rate right-of-center governments more highly relative to a survey, while a negative coefficient indicates a bias towards left-of-center governments. Looking across the columns of Table 6, we find only one source which appears to have a consistent ideological bias, with the Heritage Foundation assigning relatively higher scores to countries with right-of-center governments than the corresponding surveys. However, it is worth emphasizing that this “ideology bias” is fairly modest in magnitude. The coefficient estimates indicate that a country with a right-of-center government would get between 7 and 10 percentile points higher than a center government. Moreover, in all cases, the ideology variable in a statistical sense explains only a trivial fraction of the difference in assessments between polls and surveys, suggesting that the importance of ideological biases in polls is quite small overall.

4.3 Margins of Error and Classifying Countries According to Governance Performance

In this subsection we illustrate the importance of taking margins of error into account when assigning countries into groups according to their level of governance. Our discussion here is motivated by the recently-announced allocation rules for the new aid program of the United States Government, the Millennium Challenge Account (MCA). On November 25, 2002 the U.S. Government released details of how it plans to allocate MCA funds towards countries that “govern justly”, “invest in people”, and “promote economic freedom”.²¹ This represents a major policy shift by a donor in moving

²⁰ For Voice and Accountability, we use GMS instead of WBS as the comparator survey. This is because the WBS questions used in VA are a bit special, capturing the extent to which firms have voice in the policymaking process, which is considerably narrower than most other polls. In contrast, GMS asks questions that more broadly correspond to the concepts covered by the polls of experts.

²¹ See Radelet (2003) for a detailed discussion of the MCA.

to an allocation criteria which places governance issues center stage, and which relies on a highly transparent and objective empirical allocation criteria.²² The proposed criteria for country eligibility draw heavily on a number of cross-country measures of the quality of governance, including the five of the six governance indicators presented here..

The proposed MCA allocation rule is designed to ensure that MCA funds will be allocated to low-income countries with relatively sound policies and institutions. A group of 74 countries that are eligible for concessional IDA lending from the World Bank, and which have per capita incomes less than US\$1,435 in 2001, will potentially be eligible for MCA funds in its first year.²³ According to the MCA eligibility rules, this set of countries will be rated according to 16 performance criteria covering three dimensions of performance: “governing justly” (6 criteria), “investing in people” (4 criteria), and “promoting economic freedom” (6 criteria). Four of the governance research indicators we have constructed (voice and accountability, government effectiveness, rule of law, and corruption) have been proposed as performance indicators under the MCA’s “governing justly” performance dimension, with the remaining two for this dimension being measures of civil liberties and political rights constructed by Freedom House. In addition, a fifth governance indicator, Regulatory Quality, is included under “promoting economic freedom”.

In order to qualify for MCA assistance, countries must (a) be in the top half of all potentially eligible countries according to the corruption rating from the governance indicators, and (b) must be in the top half of all potentially eligible countries on at least half of each of the performance criteria under each of the three dimensions of performance. This rule is designed to ensure that resources are channeled towards

²² However, the MCA is of course not the only organization to make explicit use of governance indicators. For example the World Bank uses its internal assessments of countries policy performance, the Country Policy and Institutional Assessment (CPIA ratings), to allocate concessional lending from its IDA facility

²³ A number of countries with per capita income greater than US\$1,435 are currently eligible for IDA under the small island economies exception, but these will not be eligible for the MCA during the first year. The group of 74 countries is based on data on IDA eligibility available at www.worldbank.org/ida and per capita gross national income in U.S. dollars in 2001 using Atlas exchange rates available in the World Bank’s World Development Indicators. There will likely be differences between this list and the official list compiled by the MCA administration due to differences in data sources and updates, etc. In the second year, eligibility will expand to all countries with per capita incomes less than US\$1,435, and in the third year to all countries with per capita incomes less than US\$2,975.

countries that are performing well in a variety of dimensions of governance, and in which corruption especially is relatively low. Given the abundant evidence of the importance of good institutions and policies for growth, development, and aid effectiveness, this type of allocation rule is certainly warranted. Moreover, an objective and monitorable set of criteria for determining MCA eligibility is highly desirable, both in terms of the process of aid allocation, and also in terms of creating clear incentives among potential recipients of this aid.

However, it is important to note that the substantial margins of error associated with governance estimates mean that it is difficult to assign many countries to a definitive performance category according to their estimated level of governance. This point applies to any of the MCA criteria, or any other indicator for that matter. Recognizing these margins of error, the MCA criteria appropriately do not require countries to pass the median hurdle on all indicators. However, the proposed MCA allocation rule in principle required countries to score in the top half of all relevant countries on the corruption indicator, which would constitute a “hard hurdle” for eligibility. While it is reasonable to subscribe to the notion that corruption should be an important factor in allocating aid, it is important to emphasize that a simple “in-or-out” rule runs the risk of misclassifying some countries precisely because margins of error are not trivial. This possibility is recognized in the MCA fact sheet, which provides for some flexibility to *de facto* ‘soften’ somewhat the “hard hurdle”.

Since we have explicit margins of error for our corruption indicator, it is possible to be explicit about the risks of misclassifying countries with a single hard hurdle such as this. This is done in Figure 3. We rank the 74 potential MCA countries according to their scores on the governance indicator in 2002 on the horizontal axis, and for each country, we show the estimated corruption rating as a black diamond and the 90% confidence intervals for each country as a vertical line, on the vertical axis. The first point to note from Figure 3 is that the 90% confidence intervals for a substantial fraction of countries include the median score (indicated as a heavy black horizontal line on the graph). In particular, of the 37 countries in the bottom half of the sample, only 11 have 90% confidence intervals that are fully below the median score, while for the remaining countries the confidence intervals include the median, indicating that there is at least a 10% probability that these countries’ scores are actually in the top half of the sample.

Similarly, only 17 out of 37 countries in the top half of the sample have confidence intervals that are fully above the median score for all countries, while for the remaining 20 countries there is a non-trivial probability that these countries' scores are in fact in the bottom half of the sample.

For the majority of countries there is a non-trivial probability that they could be mistakenly classified in the bottom half of the sample. To illustrate this more precisely, for each country our methodology allows us to calculate the probability that a country's true unobserved level of governance falls in the top half of the sample. These probabilities are indicated as squares, triangles and circles in Figure 3. Not surprisingly for the worst-rated countries, the probability they could fall in the top half of the sample is close to zero. Similarly, the best rated countries almost certainly belong in the top half. However, there is a large intermediate range of countries where there is a non-trivial probability that they belong in either the top or bottom half of the sample, for example ranging from around 0.25 to 0.75. Borrowing colors from a traffic light, we refer to the first group of squares as red (less than 25% chance that they are mistakenly classified in the bottom half), the second group of circles as green (more than 75% chance that they actually belong in the top half), and the intermediate group of triangles as yellow.

This "traffic light" approach highlights the challenge of assigning countries to performance categories – particularly the substantial number that fall in the "yellow light" category. While the 27 "green light countries" most likely belong in the top half, and the 22 "red light" countries in the bottom half, the 25 intermediate "yellow light" countries are a more difficult case. These countries are difficult to assign to categories simply because the available cross-country data is not sufficiently informative and/or there is disagreement between the underlying sources. The difficulty of assigning the "yellow light" countries to either the "green" or "red" categories points to the importance of relying on additional sources of information on which to base MCA eligibility decisions for this group. This also underscores the need for a certain degree of flexibility in the MCA allocation rule, and importantly, that this flexibility should be symmetric. Not only should countries that "barely miss the list of better performers" be given special consideration as currently proposed in the MCA fact sheet, but in addition countries that barely make the list of better performers should also merit further scrutiny.

In summary, the discussion of this section emphasizes the importance of caution when using governance indicators to classify countries into groups. Classifications based on individual indicators, or even on a single aggregate indicator, inevitably run the risk of misclassifying countries due to the margins of error inherent in all indicators. To reduce the risk of misclassification, it is important to look at a variety of indicators and additional sources of data, especially for borderline cases.

4.4 Margins of Error Are Not Unique to Subjective Data

We have argued that one of the strengths of the governance indicators reported in this paper is that we are able to construct explicit margins of error associated with our estimates of governance for each country. However it is worth emphasizing that these margins of error are not unique to subjective or perceptions-based measures of governance, but are also present -- if not explicitly noted -- in most other measures of institutional quality, or any other socioeconomic indicator for that matter. One need only consider the range of "preliminary" estimates of basic objective variables such as real GDP growth produced in industrial countries with high-quality statistical systems to realize that measurement error in objective data is in fact pervasive and should be taken seriously.²⁴

Consider for example the recent interest in constructing objective measures of governance that do not exclusively rely on perceptions-based data sources as we do, but rather on objective and quantifiable data. Several of these are described in Knack and Kugler (2002). They argue that variables such as the waiting time required to obtain a telephone line, and the number of telephone faults can serve as proxies for public administrative capacity. The reliance of the government on trade taxes can serve as a proxy for the (in)ability of the government to broaden its tax base. The volatility in budgetary expenditure shares, and similarly, the volatility of revenue shares, are indicative of a volatile and unpredictable policy environment. They also draw on a number of other measures of institutional quality pre-existing in the literature. Clague, Keefer, Knack and Olson (1996) argue that the fraction of currency in circulation that is

²⁴ For example, in mid-2002, the U.S. Bureau of Economic Analysis had produced three estimates of GDP growth for the fourth quarter of 2001, ranging from 0.2% to 1.7% (Wirtz 2002). Explicit consideration of margins of error in national accounts data dates back at least to Kuznets (1941).

held in the banking system is a good proxy of the extent to which individuals in a country can be confident that their property rights are protected. Finally, in a series of papers, Djankov et al. (2002, 2003) compile cross-country data on the number of administrative procedures required to start a business, and the number of legal procedures required to collect an unpaid debt. These measures capture the complexity of the regulatory and legal environment.

While most of these measures can in principle provide an accurate measure of the specific underlying concept to which they refer, their usefulness as a measure of broader notions of governance depends on the extent to which the specific concept they are measuring corresponds to these broader ideas of governance. For example, the number of procedures required to start a business may not be a good indicator of the complexity or burden of regulation in other areas. Similarly, the willingness of individuals to hold currency in banks reflect their confidence in a very particular set of property rights (vis-à-vis banks, and banks vis-à-vis the government), but may not necessarily capture other dimensions of property rights protection, such as confidence in the police and judicial system to uphold private property rights.

This is of course not surprising, nor should it be considered a drawback of such measures -- all of which necessarily are imperfect proxies for broader notions of governance. However, it does mean that one should consider seriously the margins of error for objective indicators as well, to the extent that these are used as proxies for broad concepts of governance such as the ones we measure using subjective data in this paper.²⁵ While these margins of error are generally not made explicit for objective indicators, a simple calculation can give a sense of their order of magnitude. Suppose that we have two noisy indicators y on a common unobserved concept of governance, g , i.e.: $y_i = g + \varepsilon_i$, $i=1,2$. Then if we normalize the variance of the unobserved measure of governance to be one, the correlation between the two observed indicators will be $\rho = \left((1 + \sigma_1^2) \cdot (1 + \sigma_2^2) \right)^{-1/2}$. Suppose that indicator 1 is one of our subjective governance indicators, for which the variance of the measurement error, σ_1^2 , is known, and that indicator 2 is one of the objective indicators described above. Then from the observed

²⁵ These margins of error should of course also reflect measurement error in the raw data on which they are based -- for example, the non-trivial measurement error in macroeconomic variables such as the money supply or the composition of public expenditures.

correlation between the two indicators, we can infer the variance of measurement error in the objective indicator, σ_2^2 .

The results of this calculation can be found in Table 7. The rows of Table 7 correspond to the various objective governance indicators discussed above. In the first two columns, we identify the objective indicator, and the subjective aggregate governance indicator which best corresponds to it. In the third column we report the correlation between the subjective and the objective indicator, using our 2002 governance indicators. The next three columns report the implied standard deviation of measurement error in the objective indicator, under three assumptions: (A) that our estimate of the standard deviation of measurement error in the subjective indicator is correct, (B) that the subjective and objective indicators have the same standard deviation of measurement error, and (C) that the standard deviation of measurement error in the subjective indicator is twice as large as that in the objective indicator. Finally in the last column we report the actual standard deviation of measurement error, computed as the average across all countries of the country-specific standard errors in our governance indicators.

The results in Table 7 are quite striking. For all indicators, and for all three sets of assumptions, the implied standard deviation of measurement error in the objective indicators is very high relative to the corresponding standard deviation of the subjective governance indicators. Under the benchmark assumption (A) which takes seriously the margins of error we have computed for our governance indicators, we find that the implied margin of error for the objective indicators is between seven and 15 times larger than that of the subjective indicators. This clearly exaggerates the difference in the precision of subjective and objective indicators because we are comparing a single objective indicator with an aggregate of several subjective measures, and as discussed we should expect aggregation to improve precision. But this is only part of the story. For the GE and RQ indicators, we have a median of six sources per country, while for RL we have a median of eight sources. This can explain why the standard deviation of measurement error of the objective sources might be $\sqrt{6} = 2.4$ to $\sqrt{8} = 2.8$ times higher than that of the corresponding subjective indicators, but still cannot explain all of the difference in the precision of the indicators that we see. Similarly, the last row in Table 6

reports the correlation of GE with an aggregate of all the objective indicators. In this case, the benefits of aggregation would be roughly comparable for the two indicators, with a median of 5 sources per country for the objective indicator and a median of 6 sources per country for GE. Nevertheless, we find that the implied standard deviation of measurement error is still four times as large for the objective indicator as it is for the subjective one.

Assumptions (B) and (C) are designed to be more favorable to the precision of the objective indicators. Assumption (B) discards the information in the margins of error that we have constructed for the subjective indicator, and simply makes the neutral assumption that the subjective and the objective indicators have the same standard deviation of measurement error. This reduces the implied standard deviation of measurement error for the objective indicator relative to the benchmark assumption (A), but it remains large at 0.6 for the composite objective indicator, and higher for the individual indicators. Assumption (C) weights things even further in favor of the objective indicators, assuming that the objective indicator is twice as precise as the subjective indicator. In this case, we continue to find very substantial estimates of the standard deviation of measurement error, on the order of 0.4 and higher for individual objective indicators.

This simple calculation underscores and helps to quantify the intuitive notion that all governance indicators, not just the subjective ones we have constructed, are subject to non-trivial margins of error, and that care should be taken in making governance comparisons based on any such measures. In addition, wherever possible, it is desirable to construct explicit margins of error to aid in these comparisons.

4.5 Global Trends in Governance

In this subsection we present the limited available evidence we have on trends in global averages of governance. This exercise is of interest because, as noted above, we have rescaled the mean of our governance indicators to be equal to zero in each period. As a result, our aggregate indicators are by construction uninformative about trends in global averages of governance, but are informative about countries' relative positions

around the average. If we want to make statements about trends in governance worldwide, we need to go back to our underlying sources of governance data.²⁶

Table 8 summarizes the information we have available on trends in world averages of the various dimensions of governance. The panels of Table 8 correspond to our six dimensions of governance. Within each panel, we report data from up to four major underlying sources that we have available to us in each of the four periods 1996, 1998, 2000, and 2002: EIU, DRI, PRS, and GCS. The first three are polls of experts that cover a large set of countries with a consistent methodology from year to year, and can therefore be expected to be informative about overall trends. GCS covers only a smaller set of countries, but it is the only survey of individuals we have that is available in all four periods, and so we also present it here. To maximize comparability across sources and over time, for the first three sources we focus on the set of countries common to these three for all periods. For GCS alone we consider only the smaller set of countries available in each period, focusing on a small number of individual survey questions that have been consistently available over time.

The first four columns present the average across all countries of each of the sources in each of the years. The underlying data have been rescaled to run from zero to one, and for each source and governance component, we report the score on the same question or average of questions that we use in the aggregate indicator. The next four columns report the standard deviation across countries for each source. The final column reports the t-statistic associated with a test of the null hypothesis that the world average score is the same in the first and in the last periods. Interestingly, a number of sources report quite substantial declines in the world averages of the six dimensions of governance. DRI in particular shows statistically significant declines in all five indicators in which it appears. PRS reports significant declines in world averages for Political stability, Rule of Law and Corruption, but an improvement in Regulatory Quality and Government Effectiveness. Among polls, EIU alone consistently does not report any significant trend. Finally the one survey in this table, GCS, reports significant deterioration in Political Stability and Government Effectiveness, and declines in Rule of Law.

²⁶ We also note that by construction the standard deviation of our aggregate governance estimates is equal to one in each period. We therefore also cannot assess whether there has been global convergence in governance based on these aggregate indicators.

It is not clear how much importance ought to be ascribed to these trends in world averages. On the one hand, these statistics represent the only information we have on trends over time, and so they should be taken seriously. On the other hand, it is striking from Table 8 the extent of disagreement among sources as to the direction of global trends – overall we see 8 averages that improve or remain the same, and 11 that decline – so we cannot infer from Table 8 that there is full agreement that governance is getting worse worldwide. However, looking only at statistically significant changes, declines in governance averages outnumber increases by ten to two (and both of the statistically significant increases are in regulatory quality; none for any of the other five governance components). For now we cautiously conclude that we certainly do not have any evidence of any significant improvement in governance worldwide, and if anything the evidence is suggestive of a deterioration, at the very least in key dimensions such as control of corruption, rule of law, political stability and government effectiveness. It can therefore be safely concluded that our (relative) governance estimates for a country do not underestimate absolute trends for such country, since there is no evidence of a worldwide improvement.

4.6 Comparisons with the Transparency International Corruption Perceptions Index

In this section we briefly contrast the methodology and results used here with those used by Transparency International in the construction of its pioneering Corruption Perceptions Index. Like the indicators presented here, the TI index is an average of ratings reported by a number of perceptions-based sources. In terms of content, the primary difference is that the TI index relies on a subset of the sources we use here, and in addition the TI index treats multiple years of data from the same source as separate sources in the aggregation procedure. In particular, the TI index is based on 10 distinct data sources, but uses between two and three years of data from some of them and treats them as separate sources, to arrive at a total of 15 components for the TI index. In contrast the control of corruption index here is based on 14 distinct sources, using only data from 2002, and without taking multiple years from the same source.

The TI index also differs in its approach to aggregation. We rely on an unobserved components model to: (i) transform individual sources into common units (this is the role of the α 's and β 's), (ii) construct an appropriately-weighted average of sources to use as the aggregate score, and (iii) produce margins of error to summarize the precision of the estimates of governance. In contrast, TI uses: (i) a percentile-matching approach to put data in common units; (ii) a simple average of rescaled scores as the estimate of corruption for each country, and (iii) a non-parametric bootstrapping procedure, discussed in more detail below, to generate measures of precision for the aggregate indicator. Details on the TI approach can be found in Lambsdorff (2002).

In terms of results, the estimates of corruption based on the two indicators are very similar, with a correlation above 0.9. However, the TI index covers only 102 countries, since it discards countries with fewer than three data sources. In contrast the control of corruption indicator covers 195 countries, or nearly twice as many. The margins of error generated by the two approaches are quite similar as well. For the TI index, the average width of a 90% confidence interval is 0.94, or 9.4% percent of their range of units from zero to 10. The average width of a 90% confidence interval of the Control of Corruption indicator constructed here 0.71, or 14% of the range from -2.5 to 2.5 of this measure. However, these figures are not comparable, since our Control of Corruption indicator covers many more countries, many of which have only one or two sources of data and hence should be expected to have somewhat larger margins of error. If we focus only on the set of 102 countries appearing in the TI index, the average width of a 90% confidence interval for the Control of Corruption indicator is 0.52, or 10% of the range of this index. This is almost identical to the margins of error in the TI index, relative to its scale of units.

However, as we discuss in the remainder of this section, this apparent similarity in the precision of the two indicators is the result of two offsetting biases in the TI methodology. On the one hand, the bootstrapping approach to calculation margins of error in the TI approach understates the margins of error, i.e. overstates the precision of estimates of corruption, for countries with relatively few sources. This would suggest that the TI margins of error should *spuriously* be smaller than those we report. On the other hand, the estimates of corruption produced by the unobserved components model (UCM) used in this paper are a precision-weighted average of individual sources, while

the TI approach is based on a simple average. Since precision-weighting improves the accuracy of the estimates of corruption, we should expect that the TI margins of error should *correctly* be larger than the ones we report.

We make these points with the help of some simple examples. In order to provide a concrete framework in which to discuss these issues, we begin by assuming that after rescaling to common units, each source of governance data k provides a noisy signal about the true underlying level of governance in country j :

$$(4.1) \quad y(j,k) = g(j) + \varepsilon(j,k)$$

This is identical to Equation (2.1), except that we have assumed that all sources have already been transformed to common units, i.e. $\alpha(k)=1$ and $\beta(k)=1$ for all sources k . As discussed above, the UCM approach treats $g(j)$ as a random variable, and estimates governance for each country as the mean of this variable *conditional on the data observed for that country*. We have seen that this conditional mean is simply a weighted average of the observed governance data for each country, with weights proportional to the precision of each source.

The TI approach takes the somewhat different interpretation that $g(j)$ is a fixed parameter to be estimated for each country. It then constructs a simple average across sources of the $y(j,k)$ for each country as its estimate of corruption. The precision of this estimate is summarized by a bootstrapping procedure. This bootstrapping procedure is based on the idea that one can approximate the range of likely values of governance for a country by looking at averages based on all possible combinations of the observed data for each country. In practice, this means taking a large number of random draws from the K data sources for each country, and constructing the average corruption score for each draw. Then the 5th and 95th percentiles of the distribution of these many sample means provides a likely range of governance estimates that is comparable to the 90% confidence intervals based on the UCM.

We can now illustrate our two main points:

1) *If the number of sources available for a country is small, the TI approach of bootstrapping will tend to produce spuriously small margins of error and hence will tend to exaggerate precision.* The reason for this downward bias in margins of error is quite intuitive. Consider the extreme case where there is only one data source available for a country. Then, it simply is not possible to construct bootstrapped margins of error, because there is only one possible combination of sources for that country. Less trivially, suppose there are only two sources for a country. This means that there are three possible values for the sample average based on different combinations of this data. The key issue is whether the range of these three averages accurately captures the true variation in the data. As long as the number of observations is small, the range of averages based on bootstrapping will tend to be smaller than the true range of variation, and so bootstrapping will tend to understate precision.

The easiest way to see this is with a simple numerical example. Suppose that the “true” governance data is generated by Equation (4.1), with $\sigma=0.5$ for all sources (roughly corresponding to a typical value that we estimate using the UCM). We then generate random data from Equation (4.1) assuming normally-distributed errors, and then calculate bootstrapped 90% confidence intervals analogous to those constructed by TI. The top panel of Figure 4 reports these as a bold line, for a range of values of the number of sources per country from 2 to 10. Superimposed on the same graph are “true” the 90% confidence intervals as well, with widths equal to $3.28 \cdot \sigma / \sqrt{K}$. The main point of this example is that the downward bias in bootstrapped standard errors is substantial when the number of sources is small. For countries with three sources (the minimum number in order to qualify for inclusion in the TI index), bootstrapped margins of error understate the “true” margins of error by about 40% (0.59 as opposed to a “true” value of 0.96). This downward bias is smaller the more sources are available for a given country, but is still not trivial. For countries with four sources, the downward bias is near 20%, and with five sources the bias is 10%. Since about 40% of the countries included in the TI index have between three and five sources, these biases apply to a substantial fraction of the TI sample. Overall, this example highlights one potential drawback of

bootstrapping, which is that it assigns spuriously high precision to countries with relatively few data sources.²⁷

2) *If there are large differences across sources in the precision with which they measure governance, a governance estimate based on a simple average of scores such as the TI index will be unnecessarily imprecise. In contrast, weighting sources by their reliability can improve the precision of the overall estimates.* Again, a simple example best illustrates this general point. If the variances of the error terms are different, a weighted average of sources will have a smaller variance than a simple average, if the weights are proportional to the precision of the sources. Suppose that the standard deviation of measurement error of source k is $\sigma(k)$. Then the minimum variance estimator of the mean level of governance for country j , $\bar{y}^*(j)$, will be:

$$(4.2) \quad \bar{y}^*(j) = \sum_{k=1}^K \left(\frac{\sigma^{-2}(k)}{\sum_{i=1}^K \sigma^{-2}(i)} \right) \cdot y(j,k)$$

Note that this weighted average is very similar to the weighted average of scores constructed by the UCM (recall Equation (2.2)).

How big are the efficiency gains of a weighted average, as opposed to the simple average of scores, $\bar{y}(j) = \frac{1}{K} \sum_{k=1}^K y(j,k)$, as constructed by TI? The answer to this question will depend on how different are the variances across sources. Some simple arithmetic shows that the ratio of the standard deviation of these two alternative estimators of governance is:

$$(4.3) \quad \sqrt{\frac{V[\bar{y}(j)]}{V[\bar{y}^*(j)]}} = \frac{1}{K^2} \cdot \sum_{i=1}^K \sum_{k=1}^K \frac{\sigma^2(i)}{\sigma^2(k)}$$

²⁷ This is not to say that bootstrapping lacks any merit. In principle, an important advantage of bootstrapping is that it will give more reliable measures of precision when the underlying data are not normally distributed. However, it is difficult to say a priori the extent to which deviations from normality will impart *systematic* biases into the margins of error generated by the UCM.

In order to get a sense for the magnitude of this quantity, we take the estimates of $\sigma(k)$ obtained from the UCM, for each of our sources in 2002. For each indicator we then calculate what the ratio of the standard deviation of the unweighted mean (calculated by TI) would be relative to the optimal weighted mean, for a hypothetical country appearing in all sources for a given indicator. The results of this are summarized in Table 9, and show that the benefits of precision-weighting are large for the set of sources we use. The variance of the simple average of scores is between 19% higher than the weighted averages (in the case of Political Instability and Violence), and 121% higher (in the case of Rule of Law). On average, constructing a precision-weighted average results in margins of error that are 75% smaller than those of a simple unweighted average. These large benefits of weighting simply reflect the fact that there are very substantial differences in the precision of the individual sources we use, and that it is optimal to assign greater weight to sources with relatively higher precision.

In summary, we have seen that there are two opposing forces in the comparison of the margins of error generated by TI, and those from the UCM that we use here. First, bootstrapped standard errors as constructed by TI will be spuriously small when the number of sources is small. Second, simple averaging as is done by TI will correctly lead to higher standard errors because the simple average is not as efficient as the weighted average constructed by the UCM.

We can now quantify the relative importance of these two factors. In order to do so, we first apply the UCM methodology to the data used by TI to construct its 2002 index for 102 countries. When we compute the average confidence range based on the UCM for these countries using the TI data and sample of 102 countries, we find an average value of 0.64, which is 37% larger than that reported by TI. It is important to note that these margins of error are also substantially larger than the margins of error of our 2002 Control of Corruption index for these countries. This is because the Control of Corruption indicator is based on a greater number of independent data sources than the TI index (14 independent sources, versus 10 for TI).

When we apply the UCM methodology to the TI set of sources, we find that the standard deviation of measurement error is quite similar across sources, ranging from

0.4 to 0.6. Calculating the efficiency loss attributable to not precision-weighting using Equation (4.3), we find that since the sources used by the TI index have such similar error variance, the unweighted average will have a standard deviation that is only 4% higher than the optimally-weighted average. This means that, for the limited sample of sources used by TI (and unlike the larger sample we use), the benefits of precision-weighting are small, and we should not expect to see sizeable differences in margins of error between the two approaches due to this factor.

In light of this, the most plausible explanation for the smaller average margins of error reported by TI is the downward bias due to bootstrapping when the number of observations is small. To illustrate this point, the bottom panel of Figure 4 compares the average width of a 90% confidence interval based on the TI methodology with that of the UCM, using the same data underlying the 2002 Corruption Perceptions Index. On the horizontal axis we show the number of data sources per country, while on the vertical axis we show the width of a 90% confidence interval, averaging across all countries with the same number of sources. For countries with very few sources, the TI method gives a much smaller standard error than the UCM. In particular, as stated, for countries with only three sources, the UCM gives an average confidence interval that is almost 40% large than that reported by TI. However, as the number of sources increases, this discrepancy gradually declines, and for a country with 12 sources the 90% confidence intervals generated by the two methodologies are almost identical.

This pattern is exactly what one would expect from the downward bias inherent in bootstrapped standard errors when the number of sources is small. In fact, comparing the top and bottom panels of Figure 4, it is striking how similar the two graphs are. With the UCM, the margins of error decline smoothly with the number of sources. In contrast, with bootstrapping the margins of error are smaller initially (reflecting the downward bias in small samples), then increase, and finally decrease again until they approach the “true” margins of error. This leads us to the conclusion that the smaller average margins of error in the TI indicator primarily reflect the downward bias due to bootstrapping for countries where the number of sources is small.

In summary, therefore, the TI Corruption Perceptions Index and the Control of Corruption indicator reported here lead to quite similar conclusions about relative levels

of corruption, for the relatively small sample of 102 countries covered by the TI-CPI index in 2002. However, applying our UCM methodology to the TI data, we find margins of error that are on average 37% larger than those reported by TI. Most of this discrepancy likely reflects a downward bias in the TI margins of error for countries with relatively few data sources, and hence an overstatement of the precision of the TI index for these countries.

5. Conclusions

In this paper we have presented substantially expanded and updated indicators of six dimensions of governance. These are based on several hundred individual measures of governance perceptions drawn from 25 sources from 18 different organizations, covering the period 1996-2002. The governance indicators for 2002 cover up to 199 countries and territories. By aggregating large numbers of individual sources, we have both expanded country coverage and also improved the precision of the aggregate indicators. Nevertheless, as we have emphasized throughout, margins of error remain substantial relative to the units in which governance is measured, and 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, as for example has been proposed for the Millennium Challenge Account eligibility criteria. In these situations, it is important to recognize the significant risks of misclassifying countries given the inherent imprecision in these indicators.

We have also discussed 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.

We intend to continue this research project on measuring cross-country differences in governance. We hope 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.

References

- Beck, Thorsten, George Clarke, Alberto Groff, Phillip Keefer, and Patrick Walsh (2001). "New Tools in Comparative Political Economy: The Database of Political Institutions." *World Bank Economic Review* 15(1):165-76.
- Cingranelli, David and David Richards (2001) "Coding Government Respect for Human Rights." Manual Version One. Binghamton, New York: State University of New York.
- Clague, Christopher, Philip Keefer, Stephen Knack, and Mancur Olson (1999). "Contract-Intensive Money: Contract Enforcement, Property Rights, and Economic Performance." *Journal of Economic Growth* 4(2):185-211.
- 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* 46(1).
- Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2002). "The Regulation of Entry." *Quarterly Journal of Economics* 117(1):1-37.
- (2003). "Courts." *Quarterly Journal of Economics*. Forthcoming.
- Efron, Bradley and Carl Morris (1971). "Limiting the Risk of Bayes and Empirical Bayes Estimators – Part 1: The Bayes Case." *Journal of the American Statistical Association* 66:807-15.
- (1972). "Limiting the Risk of Bayes and Empirical Bayes Estimators – Part 2: The Empirical Bayes Case." *Journal of the American Statistical Association* 67:130-39.
- Goldberger, Arthur (1972). "Maximum Likelihood Estimation of Regressions Containing Unobservable Independent Variables." *International Economic Review* 13(1):1-15.
- Golden, Miriam and Lucio Picci (2003). "Proposal for a New Measure of Corruption, and Tests Using Italian Data." Manuscript, University of Bologna.
- Hellman, Joel and Daniel Kaufmann (2003). "The Inequality of Influence." In *Trust in Transition*. J. Kornai and S. Rose-Ackerman, eds.
- Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton (1999a). "Aggregating Governance Indicators." World Bank Policy Research Working Paper No. 2195, Washington, D.C.
<http://www.worldbank.org/wbi/governance/pubs/aggindicators.html>.
- (1999b). "Governance Matters." World Bank Policy Research Working Paper No. 2196, Washington, D.C.
<http://www.worldbank.org/wbi/governance/pubs/govmatters.html>.

- (2002). “Governance Matters II – Updated Indicators for 2000/01.” World Bank Policy Research Working Paper No. 2772, Washington, D.C.
<http://www.worldbank.org/wbi/governance/pubs/govmatters3.html>.
- Kaufmann, Daniel and Aart Kraay (2002). “Growth Without Governance.” World Bank Policy Research Working Paper No. 2928, Washington, D.C.
<http://www.worldbank.org/wbi/governance/pubs/growthgov.html>.
- Knack, Steve and Mark Kugler (2002). “Constructing an Index of Objective Indicators of Good Governance.” Manuscript. Washington, D.C.: The World Bank.
- Kuznets, Simon (1941). *National Income and Its Composition 1919-1938*. New York: National Bureau of Economic Research.
- Lambsdorff, Johann (2002). “Background Document to 2002 Corruption Perceptions Index”. http://www.gwdg.de/~uwwv/downloads/2002_CPI_FD.pdf.
- Radelet, Steven (2003). *Challenging Foreign Aid: A Policymaker’s Guide to the Millennium Challenge Account*. Washington, D.C.: Center for Global Development.
- United States Government (2002). Fact Sheet: The Millennium Challenge Account. November 25.
- Wirtz, Ronald (2002). “Gross Domestic Product: Understanding News from Noise.” *The Region* (Federal Reserve Bank of Minneapolis): June.
<http://minneapolisfed.org/pubs/region/02-06/gdp.cfm>.

Table 1: Sources of Governance Data

<u>Source</u>	<u>Publication</u>	<u>Code</u>	<u>Type 1/</u>	<u>Country Coverage 2/</u>	<u>Repre- sentative</u>	<u>1996</u>	<u>1998</u>	<u>2000</u>	<u>2002</u>
Afrobarometer	Afrobarometer Survey	AFR	S	12					x
Business Environment Risk Intelligence	Business Risk Service	BRI	P	50		x	x	x	x
Business Environment Risk Intelligence	Qualitative Risk Measure in Foreign Lending	QLM	P	115	x	x	x	x	x
Columbia University	State Capacity Project	CDU	P	98	x			x	x
Economist Intelligence Unit	Country Risk Service	EIU	P	115	x	x	x	x	x
European Bank for Reconstruction and Development	Transition Report	EBR	P	26		x	x	x	x
Freedom House	Nations in Transition	FHT	P	27		x	x	x	x
Freedom House	Freedom in the World	FRH	P	192	x	x	x	x	x
Gallup International	Gallup Millennium Survey	GMS	S	60					
Gallup International	50th Anniversary Survey	GALLUP	S	44			x		
Gallup International	Voice of the People Survey	GAL	S	46					x
Heritage Foundation/Wall Street Journal	Economic Freedom Index	HER	P	161	x	x	x	x	x
Institute for Management and Development	World Competitiveness Yearbook	WCY	S	49		x	x	x	x
Latinobarometro	Latinobarometro Surveys	LBO	S	17		x		x	x
Political Risk Services	International Country Risk Guide	PRS	P	140	x	x	x	x	x
PriceWaterhouseCoopers	Opacity Index	PWC	S	35				x	
Reporters Without Borders	Reporters Without Borders	RSF	P	138	x				x
Global Insight's DRI McGraw-Hill	Country Risk Review	DRI	P	111	x	x	x	x	x
State Department / Amnesty International	Human Rights Report	HUM	P	159	x	x	x	x	x
World Bank	Business Enterprise Environment Survey	BPS	S	18				x	x
World Bank	World Business Environment Survey	WBS	S	81	x		x	x	
World Bank	Country Policy and Institutional Assessments	CPIA	P	136		x	x	x	x
World Economic Forum	Global Competitiveness Report	GCS	S	75		x	x	x	x
World Economic Forum	Africa Competitiveness Report	GCSA	S	23			x		
World Markets Research Center	World Markets Online	WMO	P	186	x				x

1/ P=Poll, S=Survey

2/ Countries included most recently available version of source

Table 2: Summary Statistics on Governance Indicators

	<i>Voice and Accountability</i>	<i>Political Stability</i>	<i>Government Effectiveness</i>	<i>Regulatory Quality</i>	<i>Rule of Law</i>	<i>Control of Corruption</i>	<i>Overall</i>
Number of Countries							
1996	192	165	180	182	167	151	173
1998	192	166	184	185	186	184	183
2000	192	166	185	186	186	185	183
2002	199	186	195	195	195	195	194
Median Number of Sources Per Country							
1996	4	4	4	4	6	4	4
1998	4	4	4	4	7	5	5
2000	5	6	5	4	8	6	6
2002	7	6	6	6	8	7	7
Proportion of Countries with Only One Data Source							
1996	0.15	0.13	0.21	0.15	0.07	0.18	0.15
1998	0.14	0.10	0.19	0.13	0.11	0.18	0.14
2000	0.14	0.06	0.18	0.13	0.10	0.17	0.13
2002	0.10	0.11	0.10	0.10	0.10	0.10	0.10
Average Standard Error							
1996	0.26	0.39	0.28	0.34	0.26	0.29	0.30
1998	0.25	0.32	0.30	0.34	0.25	0.25	0.28
2000	0.25	0.33	0.27	0.35	0.22	0.26	0.28
2002	0.21	0.27	0.22	0.22	0.19	0.21	0.22

Notes: This table reports the indicated summary statistics for the six governance indicators in each of the four periods. "Overall" refers to a simple average of the corresponding row of the table.

Table 3: Weights Used to Aggregate Governance Indicators

	Voice and Accountability				Political Stability				Government Effectiveness				Regulatory Quality				Rule of Law				Control of Corruption			
	1996	1998	2000	2002	1996	1998	2000	2002	1996	1998	2000	2002	1996	1998	2000	2002	1996	1998	2000	2002	1996	1998	2000	2002
Representative Sources																								
cdu	0.03	0.04	0.07	0.09	0.07	0.05	0.03	0.03	0.08	0.06
dri	0.25	0.17	0.19	0.14	0.08	0.07	0.12	0.07	0.09	0.08	0.23	0.03	0.03	0.06	0.08	0.09	0.09	0.05	0.09	0.06
eiu	0.31	0.12	0.09	0.18	0.13	0.14	0.23	0.21	0.19	0.18	0.20	0.21	0.23	0.21	0.34	0.08	0.06	0.20	0.24	0.07	0.11	0.12
frh	0.14	0.22	0.14	0.12
her	0.10	0.05	0.09	0.06	0.05	0.08	0.07	0.08
hum	0.05	0.10	0.06	0.04	0.10	0.10	0.08	0.04	0.02	0.02	0.02	0.01
prs	0.09	0.14	0.14	0.07	0.09	0.25	0.07	0.07	0.05	0.02	0.07	0.05	0.03	0.03	0.06	0.09	0.03	0.04	0.03	0.02	0.05	0.01	0.03	0.03
rsf	0.02
wbs	..	0.01	0.00	0.07	0.01	0.06	0.03	0.00	0.01	0.05	0.02	0.05	0.07	..
wmo	0.06	0.17	0.13	0.26	0.11	0.09
Non-Representative Sources																								
afr	0.01	0.01	0.02
bps	0.01	0.00	0.00	0.08	0.01
bri	0.22	0.10	0.12	0.11	0.09	0.06	0.08	0.06	0.08	0.06	0.09	0.06	0.03	0.01	0.01	0.01
ebr	0.07	0.23	0.06	0.11
fht	0.30	0.37	0.49	0.39	0.31	0.15	0.12	0.24	0.16	0.17	0.27	0.13	..	0.18	0.23	0.22
gal	0.01	0.01
gcs	0.02	0.13	0.07	0.04	0.05	0.06	0.09	0.09	0.06	0.14	0.19	0.09	0.04	0.07	0.10	0.14	0.07	0.09	0.12	0.10	0.06
gcsa	0.04	0.15	0.15	0.00	0.17
gms	0.02	0.01	0.00	0.00	0.04	0.02	..
lob	0.08	0.01	0.03	0.00	0.01	0.00	..	0.00	0.06
opf
pia	0.13	0.15	0.12	0.06	0.08	0.08	0.21	0.07	..	0.11	0.10	0.04	..	0.11	0.08	0.04
qlm	0.11	0.12	..	0.07	0.10	0.10	..	0.11
wcy	0.01	0.00	0.01	0.01	0.13	0.06	0.05	0.04	0.06	0.04	0.21	0.16	0.20	0.10	0.10	0.08	0.08	0.06	0.36	0.07	0.09	0.08

Notes: This table reports the weights applied to each source in constructing the aggregate governance indicators. The weights correspond to those that would be applied for a hypothetical country appearing in all of the available sources for that indicator. The weights are proportional to the inverse of the variance of the estimate of measurement error for each source, and the exact formula for these weights is given in the text in Section 3.2. For a country appearing in fewer sources, the relative weights applied to each source will be the same as the relative weights implicit in this table.

Table 4: Large Changes in Governance Estimates, 2000-2002

	<u>Governance Score</u>		<u>Sources Available in Both Periods</u>			<u>Agree/(Agree+Disagree)</u>	<u>Changes in Sources</u>	
	<u>2002</u>	<u>2000</u>	<u>Agree</u>	<u>No Change</u>	<u>Disagree</u>		<u>Added</u>	<u>Dropped</u>
Voice and Accountability								
Sierra Leone	-0.57	-1.36	2	2	0	1.00	2	0
Political Stability								
Sri Lanka	-0.90	-1.80	3	0	3	0.50	1	0
Namibia	0.46	-0.72	3	0	1	0.75	3	1
Argentina	-0.74	0.46	7	0	1	0.88	2	1
Cote d'Ivoire	-2.04	-0.88	3	0	1	0.75	2	1
Georgia	-1.90	-0.85	3	0	0	1.00	1	2
Israel	-1.35	-0.47	4	0	3	0.57	1	0
Kirghizstan	-1.21	-0.03	1	1	0	1.00	1	1
United States	0.34	1.26	5	1	1	0.83	1	2
Venezuela	-1.20	-0.48	4	1	2	0.67	3	1
Government Effectiveness								
Dominica	0.32	-0.86	1	0	0	1.00	1	0
Argentina	-0.49	0.30	7	1	0	1.00	1	1
Egypt	-0.32	0.35	4	2	1	0.80	1	1
Gambia	-0.81	0.25	1	0	1	0.50	2	0
Tunisia	0.65	1.32	4	0	1	0.80	1	1
Regulatory Quality								
Afghanistan	-1.82	-3.57	1	0	0	1.00	1	0
Belarus	-1.67	-2.65	1	1	2	0.33	3	1
Iraq	-2.31	-3.36	0	0	3	0.00	2	0
Moldova	0.80	0.14	5	0	0	1.00	3	1
Russia	-0.30	-1.55	6	0	1	0.86	3	1
Zaire	-1.77	-2.87	3	0	0	1.00	1	1
Argentina	-0.84	0.44	5	0	1	0.83	2	1
Bangladesh	-1.05	-0.02	2	2	0	1.00	3	1
Cameroon	-0.88	0.12	3	1	0	1.00	2	1
El Salvador	0.04	1.12	2	1	1	0.67	2	1
Zambia	-0.60	0.43	3	0	1	0.75	2	1
Rule of Law								
Samoa	0.94	-0.14	1	0	0	1.00	1	0
Argentina	-0.73	0.18	8	1	2	0.80	2	4
Cote d'Ivoire	-1.21	-0.53	2	3	1	0.67	2	1
Georgia	-1.17	-0.56	4	3	0	1.00	2	2
Namibia	0.45	1.21	2	3	1	0.67	2	1
Zimbabwe	-1.33	-0.73	6	2	0	1.00	1	0
Control of Corruption								
Madagascar	0.14	-0.80	0	1	1	0.00	1	2
Belarus	-0.78	-0.07	4	1	0	1.00	1	2
Malawi	-0.91	-0.22	2	2	0	1.00	2	2
Namibia	0.21	1.16	3	0	1	0.75	4	1
West Bank/Gaza	-0.99	0.76	0	0	0	..	1	1
<i>All Countries</i>								
Average			3.11	0.78	0.81	0.79	1.78	1.00

Notes: This table summarizes the sources of changes in estimates of governance between 2000 and 2002, for each country where the 90% confidence intervals for the level of governance in the two periods do not overlap. The first two columns report the governance scores in 2000 and 2002. The next three columns report the number of individual sources which agree/disagree with the direction of change of the aggregate indicator, and the number of sources registering no change. The final two columns report the number of sources added and dropped for each country between the two periods.

**Table 5: Agreement Among Sources over Direction of Change
in Governance Estimates 2000-2002, All Countries**

	<u>Agree</u>	<u>No Change</u>	<u>Disagree</u>	<u>Agree/(Agree+Disagree)</u>
Voice and Accountability	2.10	0.59	1.27	0.64
Political Stability	2.83	0.30	1.58	0.66
Government Effectiveness	2.26	0.42	1.55	0.62
Regulatory Quality	2.00	0.22	1.47	0.57
Rule of Law	2.58	2.30	1.63	0.62
Control of Corruption	1.96	1.75	1.32	0.62
Average	2.29	0.93	1.47	0.62

Notes: This table summarizes the sources of changes in estimates of governance between 2000 and 2002, for all countries with data in both periods. The first three columns report the number of individual sources which agree/disagree with the direction of change of the aggregate indicator, and the number of sources registering no change.

Table 6: Ideology Regressions

	<u>PRS</u>	<u>PIA</u>	<u>EIU</u>	<u>DRI</u>	<u>CDU</u>	<u>BRI</u>	<u>QLM</u>	<u>HUM</u>	<u>EBR</u>	<u>HER</u>	<u>FRH</u>
Voice and Accountability											
ideology	-2.78		-1.64		-1.72			3.67			-0.83
	0.59		0.46		0.27			0.68			0.23
Observations	44		43		28			46			46
Adjusted R-Sq	-0.01		-0.02		-0.04			-0.01			-0.02
Political Stability											
ideology	12.37		8.86	8.54	4.97	3.15		12.11			
	2.68**		1.80*	1.87*	0.93	0.61		2.52**			
Observations	52		51	46	42	25		56			
Adjusted R-Sq	0.1		0.04	0.05	-0.01	-0.02		0.09			
Government Effectiveness											
ideology	-1.84	-0.66	-2.38	1.86	-7.12	1.64					
	0.64	0.16	0.68	0.48	1.90*	0.25					
Observations	52	47	51	46	42	25					
Adjusted R-Sq	-0.01	-0.02	-0.01	-0.02	0.05	-0.04					
Regulatory Quality											
ideology	8.05	13.3		3.22					6.55	10.24	
	1.57	2.08**		0.45					0.88	1.77*	
Observations	52	47		46					15	56	
Adjusted R-Sq	0.02	0.07		-0.02					-0.01	0.04	
Rule of Law											
ideology	1.52	3.39	5.61	5.67	4.68	7.32	6.47	5.32		7.42	
	0.41	0.73	1.65	1.46	1.21	1.65	1.63	1.19		1.91*	
Observations	52	47	51	46	42	25	49	56		56	
Adjusted R-Sq	-0.02	-0.01	0.03	0.02	0.01	0.05	0.03	0.01		0.05	
Control of Corruption											
ideology	3.05	1.4	0.31	0.57	-2.21	2.83	1.84				
	0.63	0.34	0.1	0.18	0.68	0.46	0.58				
Observations	52	47	51	46	42	25	49				
Adjusted R-Sq	-0.01	-0.02	-0.02	-0.02	-0.01	-0.03	-0.01				

Notes: This table reports the results of a series of cross-country regressions of the difference in percentile rank between each poll of experts and the corresponding question from the World Business Environment Survey, on an indicator variable taking the value 1 if the government of a country is left-of-center, 2 if it is center, and 3 if it is right-of-center. For Voice and Accountability we use a question from the Gallup Millennium Survey instead of WBS. Percentile ranks are on a scale from 0 to 100 and are based on the sample of countries common to each pair of sources. All data refer to 2000. The table reports the slope coefficient, t-statistic, number of observations, and adjusted R-squared. Intercepts are not reported.

Table 7: Imputed Margins of Error for Objective Governance Indicators

<u>Objective Indicator</u>	<u>Corresponding Subjective Indicator</u>	<u>Absolute Value of Correlation</u>	<u>Implied Margin of Error for Objective Indicator</u>			<u>Actual Margin of Error for Subjective Indicator</u>
			<u>(A)</u>	<u>(B)</u>	<u>(C)</u>	
Telephone Wait Time	GE	0.56	1.43	0.88	0.58	0.21
Phone Faults	GE	0.32	2.92	1.47	1.00	0.21
Trade Tax Revenue	GE	0.50	1.68	1.00	0.67	0.21
Budgetary Volatility	GE	0.50	1.68	1.00	0.67	0.21
Revenue Source Volatility	GE	0.49	1.71	1.01	0.67	0.21
Contract Intensive Money	RL	0.57	1.39	0.86	0.57	0.19
Contract Enforcement	RL	0.40	2.25	1.22	0.82	0.19
Regulation of Entry	RQ	0.50	1.67	1.00	0.66	0.22
Aggregate Objective Indicator	GE	0.73	0.88	0.60	0.39	0.21

Notes: This table reports the margins of error for objective indicators implied by the observed correlation between objective and subjective indicators, as discussed in Section 4.3 in the text.

Table 8: Global Trends in Governance, Selected Sources

	# Countries	World Average				Std. Dev. Across Countries				t-statistic for mean difference last year - first year
		1996	1998	2000	2002	1996	1998	2000	2002	
Voice and Accountability										
DRI	112
EIU	112	0.41	0.42	0.42	0.46	0.30	0.32	0.31	0.28	1.3
PRS	112	0.67	0.66	0.66	0.66	0.23	0.25	0.25	0.25	-0.3
GCS
Political Stability										
DRI	102	0.82	0.81	0.74	0.70	0.18	0.18	0.24	0.27	-4.1
EIU	102	0.55	0.53	0.58	0.55	0.29	0.29	0.30	0.28	0.1
PRS	102	0.80	0.75	0.74	0.76	0.14	0.19	0.17	0.13	-2.0
GCS
Government Effectiveness										
DRI	102	0.59	0.58	0.50	0.48	0.27	0.26	0.30	0.30	-2.9
EIU	102	0.41	0.47	0.46	0.41	0.30	0.24	0.24	0.30	0.0
PRS	102	0.63	0.67	0.59	0.67	0.24	0.12	0.10	0.15	1.5
GCS	51	0.45	0.48	0.42	0.31	0.14	0.15	0.11	0.13	-5.2
Regulatory Quality										
DRI	106	0.83	0.84	0.79	0.76	0.14	0.14	0.18	0.20	-3.0
EIU	106
PRS	106	..	0.63	0.60	0.76	..	0.20	0.22	0.21	4.6
GCS	51	0.51	0.58	0.59	0.58	0.14	0.15	0.15	0.15	2.5
Rule of Law										
DRI	102	0.73	0.73	0.67	0.65	0.20	0.20	0.23	0.24	-2.6
EIU	102	0.50	0.53	0.51	0.54	0.27	0.29	0.30	0.27	1.1
PRS	102	0.77	0.67	0.68	0.66	0.20	0.24	0.22	0.23	-3.8
GCS	51	..	0.67	0.64	0.59	..	0.21	0.23	0.24	-1.7
Control of Corruption										
DRI	102	0.61	0.60	0.54	0.53	0.25	0.26	0.30	0.31	-1.8
EIU	102	0.37	0.37	0.36	0.38	0.31	0.33	0.32	0.32	0.2
PRS	102	0.62	0.52	0.48	0.42	0.20	0.22	0.22	0.20	-7.2
GCS	51	0.56	0.57	0.59	0.57	0.13	0.14	0.11	0.14	0.3

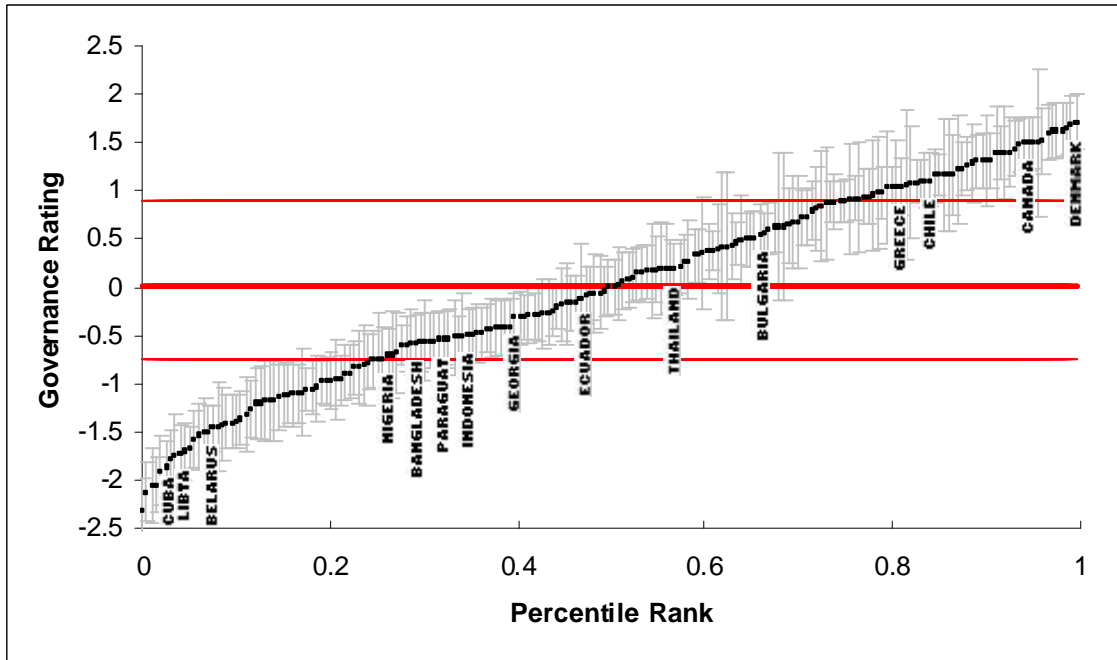
Notes: This table reports trends in cross-country averages of selected components of our governance indicators. For EIU, DRI and PRS, we restrict the sample of countries to those that appear in all three sources in all four periods, in order to ensure comparability over time and across indicators. For GCS we report the average across countries of selected individual questions that are available in all four rounds of the GCS we use, and again restrict attention to those countries available in all four periods to ensure comparability over time. Columns 2-5 contain the averages across countries, and columns 6-9 contain the corresponding standard deviations. The final column reports the t-statistic associated with a simple test for equality of global averages in the first and last available periods.

Table 9 – Efficiency Gain of Precision-Weighting

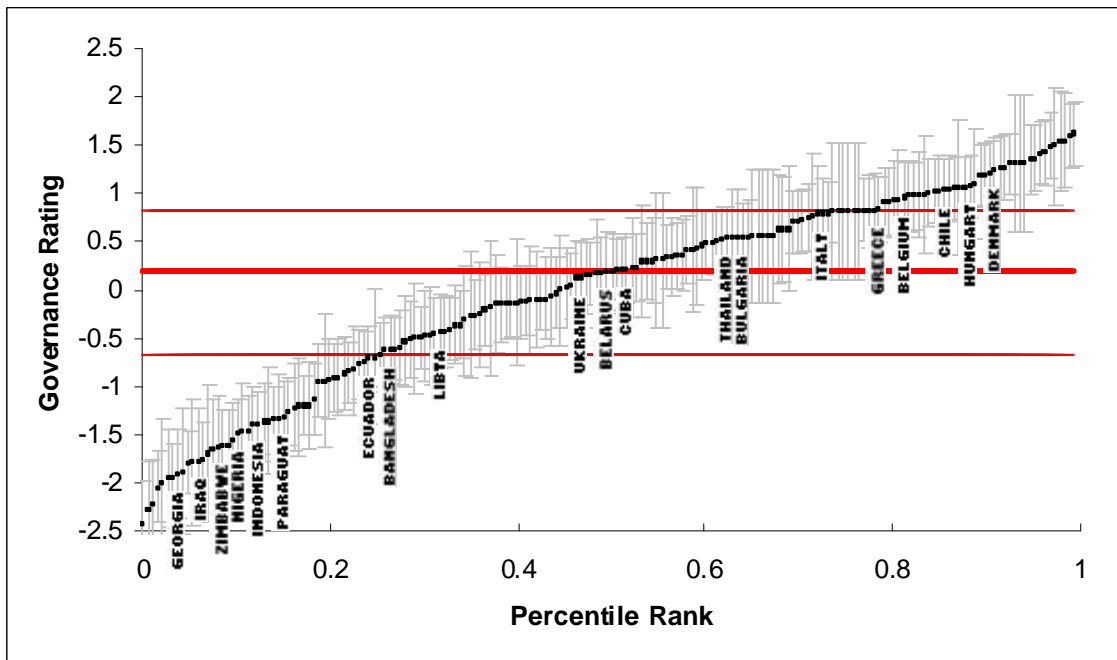
	<i>Ratio of Standard Deviation of Unweighted Mean to Precision-Weighted Mean</i>
Voice and Accountability	1.92
Political Instability and Violence	1.19
Government Effectiveness	1.69
Regulatory Quality	2.02
Rule of Law	2.21
Control of Corruption	1.45
Average	1.75

Figure 1: Estimates of Governance, 2002

Voice and Accountability



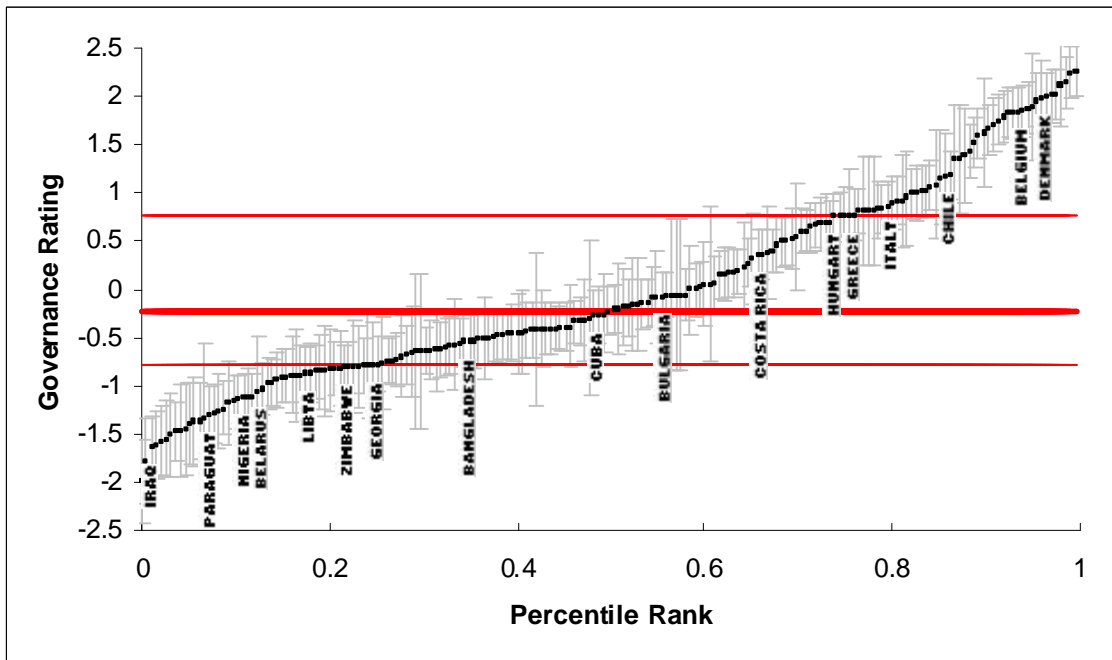
Political Stability and Absence of Violence



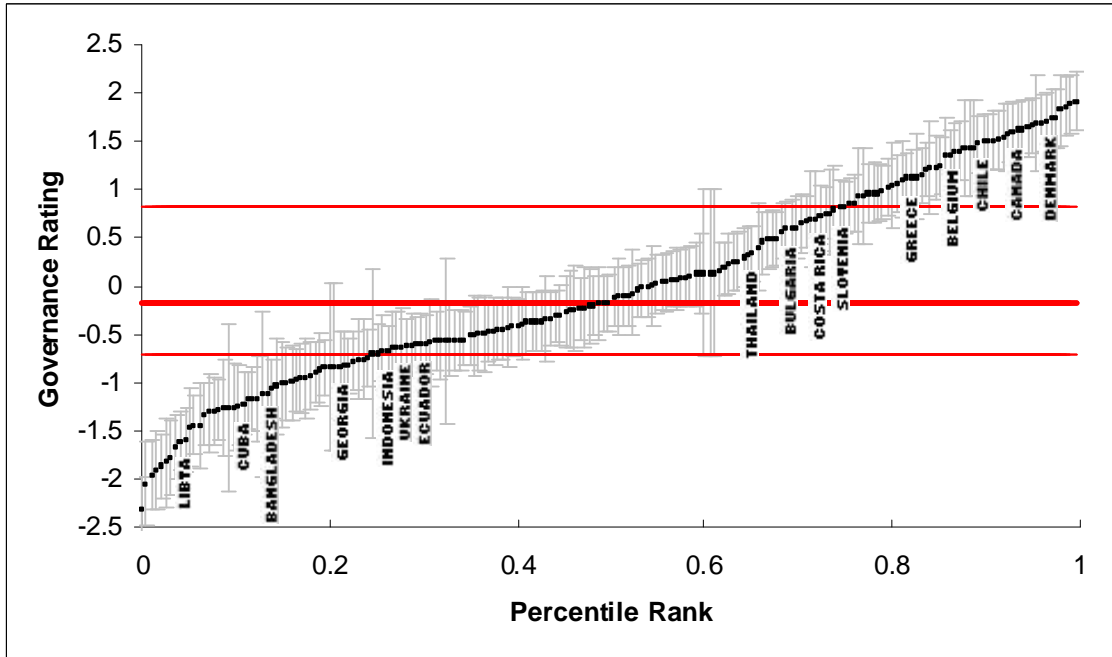
Note: This graph shows estimates of the indicated dimension of governance (on the vertical axis) for all countries graphed against each country's percentile rank (on the horizontal axis) for 2002. The vertical bars show the statistically-likely range of values of governance for each country, with the 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 1, Cont'd: Estimates of Governance, 2002

Government Effectiveness

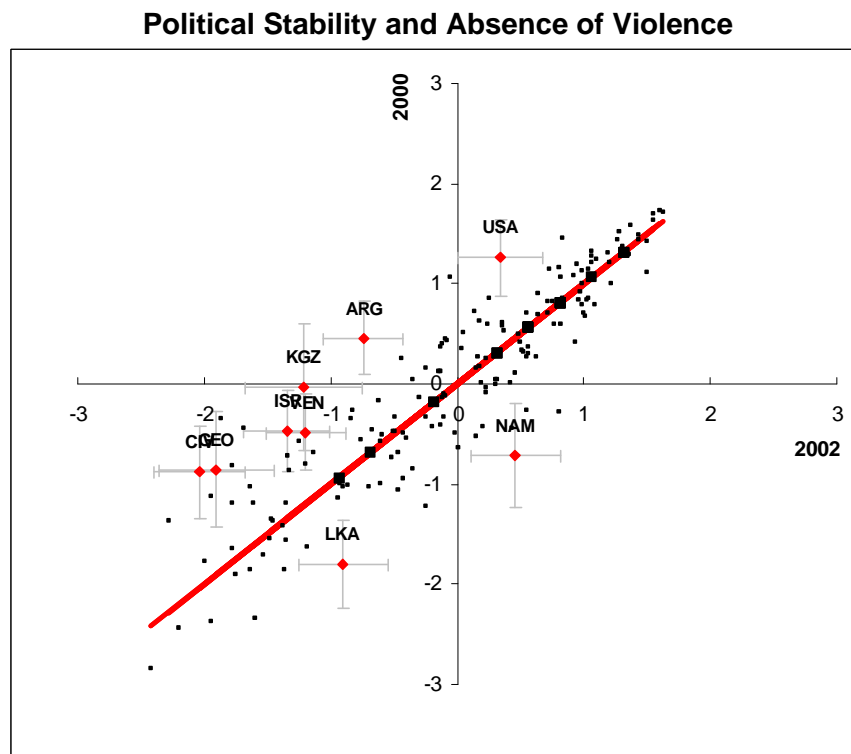
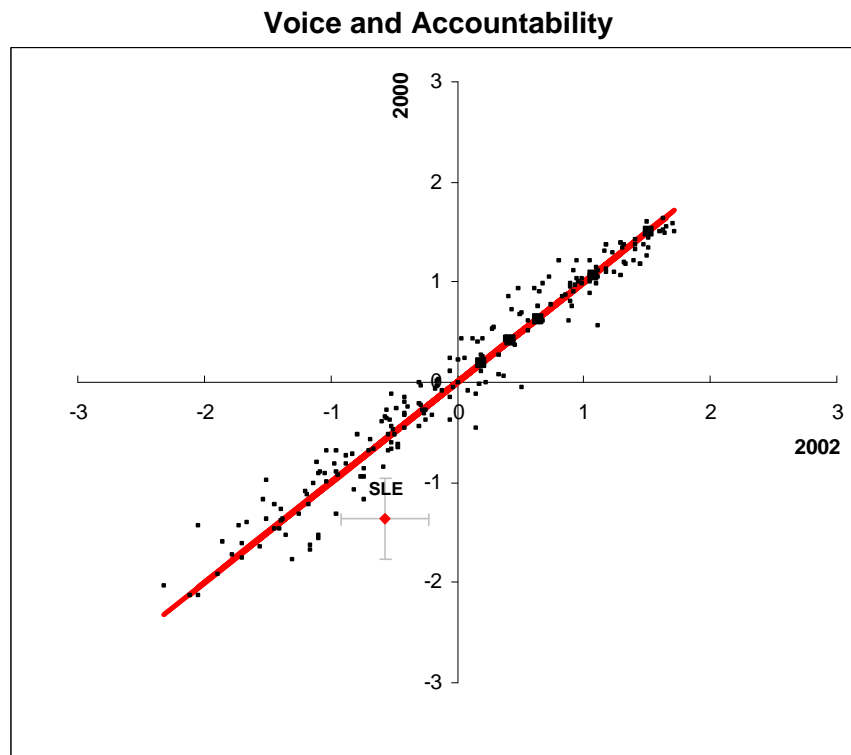


Regulatory Quality



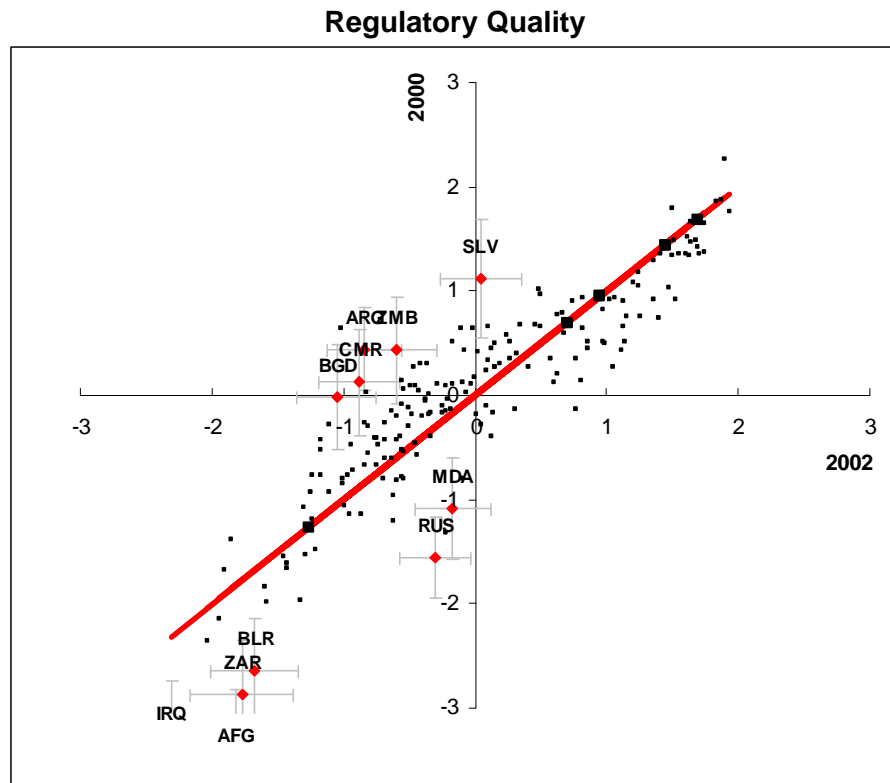
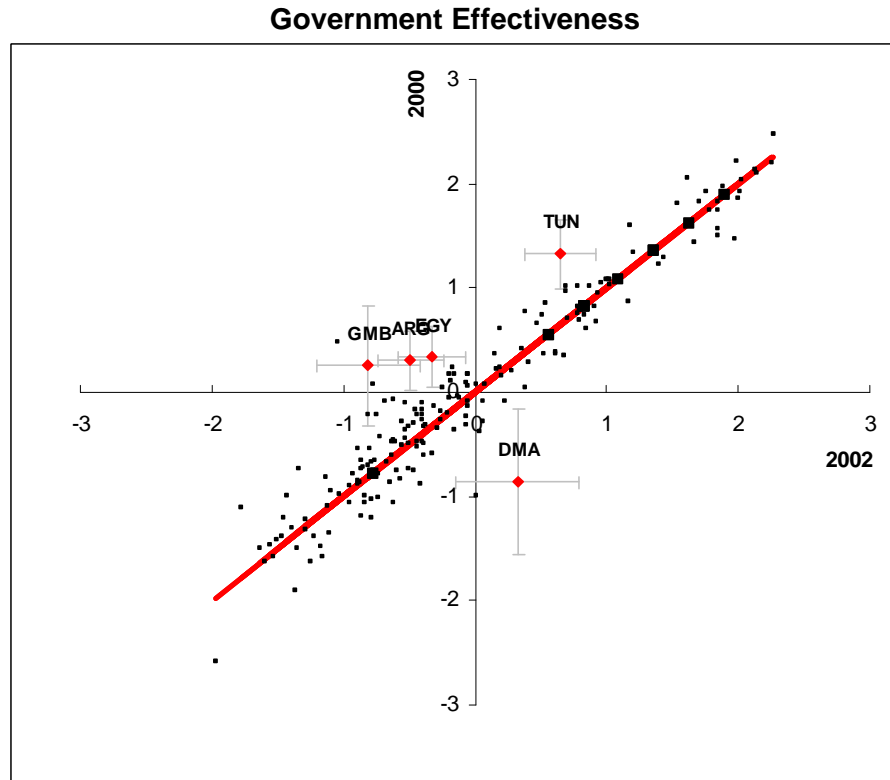
Note: This graph shows estimates of the indicated dimension of governance (on the vertical axis) for all countries graphed against each country's percentile rank (on the horizontal axis) for 2002. The vertical bars show the statistically-likely range of values of governance for each country, with the 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: Changes Over Time in Governance Estimates, 2000-2002



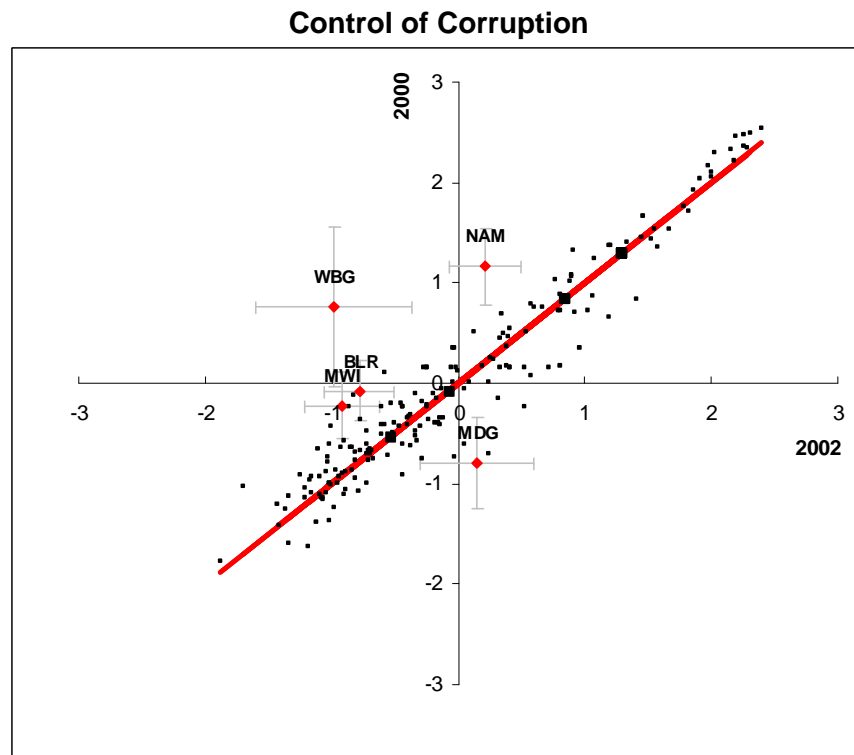
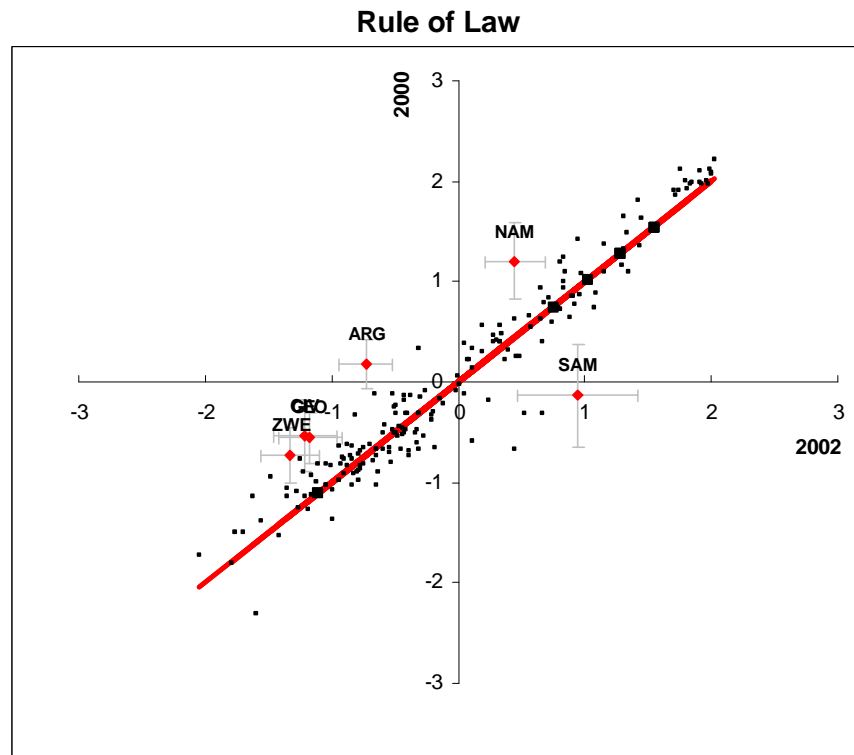
Note: This graph plots estimates of the indicated dimension of governance in 2000 (on the vertical axis) against its value in 2002 (on the horizontal axis). Countries where the 90% confidence intervals in the two periods do not overlap are highlighted, and the corresponding confidence intervals in the 2000 and 2002 are indicated as vertical and horizontal bars. As emphasized in the text, levels and changes over time in estimates of governance are subject to significant margins of error, and in no way reflect the official view of the World Bank, its Executive Directors, or the countries they represent.

Figure 2, cont'd: Changes Over Time in Governance Estimates, 2000-2002



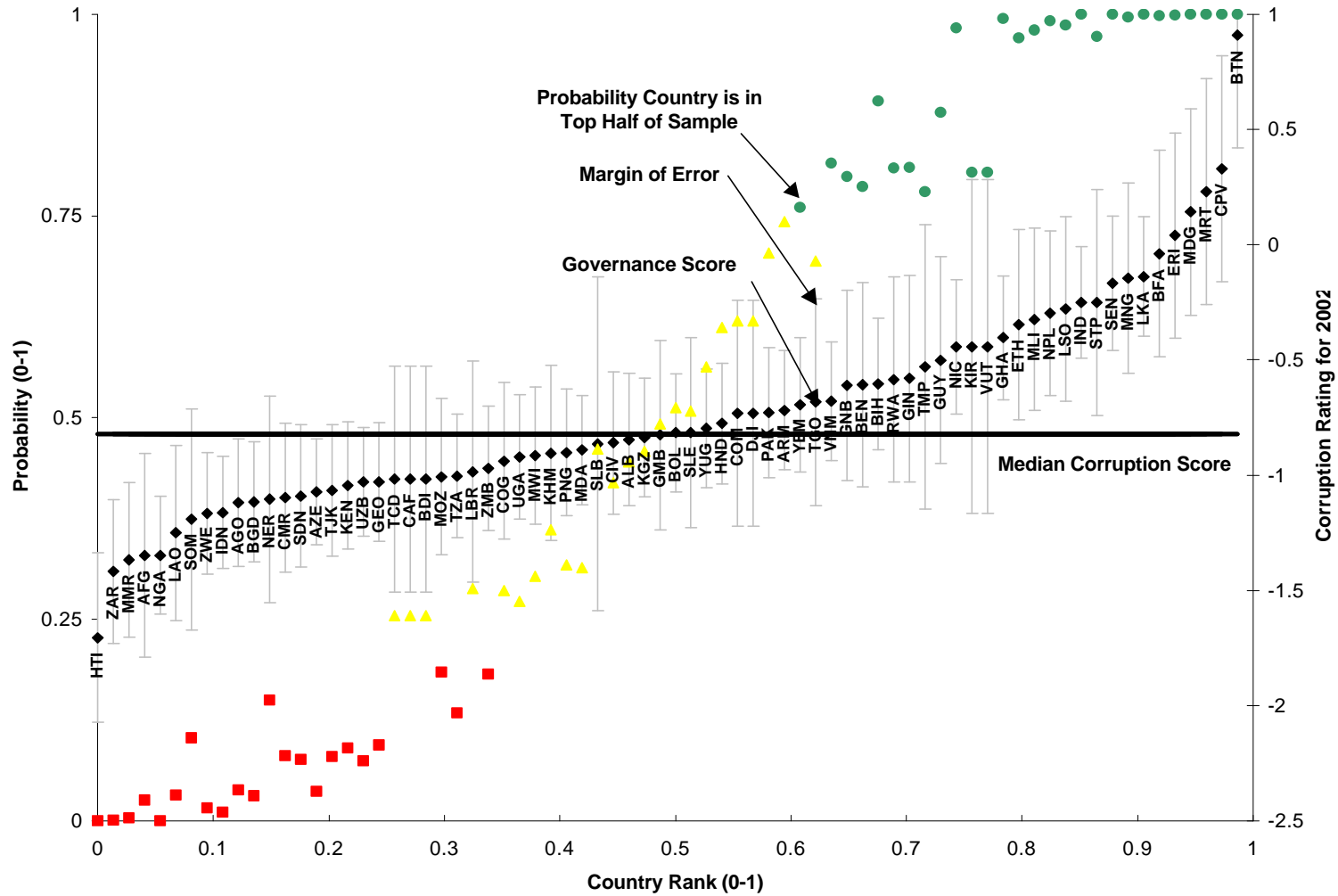
Note: This graph plots estimates of the indicated dimension of governance in 2000 (on the vertical axis) against its value in 2002 (on the horizontal axis). Countries where the 90% confidence intervals in the two periods do not overlap are highlighted, and the corresponding confidence intervals in the 2000 and 2002 are indicated as vertical and horizontal bars. As emphasized in the text, levels and changes over time in estimates of governance are subject to significant margins of error, and in no way reflect the official view of the World Bank, its Executive Directors, or the countries they represent.

Figure 2, cont'd: Changes Over Time in Governance Estimates, 2000-2002



Note: This graph plots estimates of the indicated dimension of governance in 2000 (on the vertical axis) against its value in 2002 (on the horizontal axis). Countries where the 90% confidence intervals in the two periods do not overlap are highlighted, and the corresponding confidence intervals in the 2000 and 2002 are indicated as vertical and horizontal bars. As emphasized in the text, levels and changes over time in estimates of governance are subject to significant margins of error, and in no way reflect the official view of the World Bank, its Executive Directors, or the countries they represent.

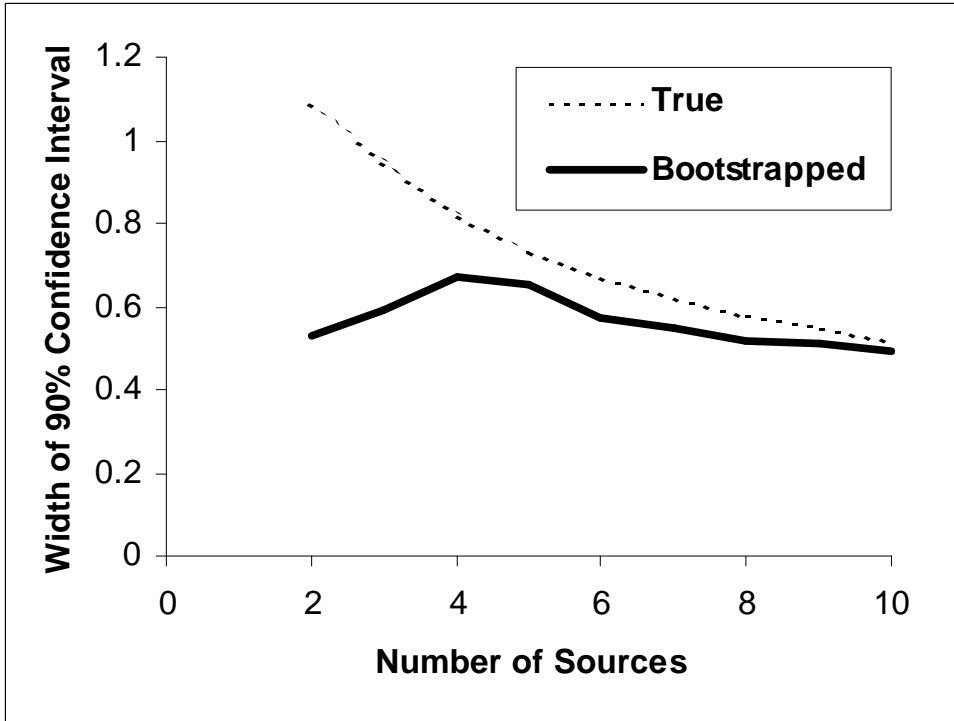
Figure 3: Using Governance Indicators to Allocate Aid



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 colors 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.

Figure 4 – Comparing TI and UCM Margins of Error

Theoretical Biases In Bootstrapped Margins of Error



Actual Differences Between TI and UCM Margins of Error

