WPS5430

Policy Research Working Paper 5430

The Worldwide Governance Indicators

Methodology and Analytical Issues

Daniel Kaufmann
Aart Kraay
Massimo Mastruzzi

The World Bank
Development Research Group
Macroeconomics and Growth Team
September 2010

Policy Research Working Paper 5430

Abstract

This paper summarizes the methodology of the Worldwide Governance Indicators (WGI) project, and related analytical issues. The WGI cover over 200 countries and territories, measuring six dimensions of governance starting in 1996: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. The aggregate indicators are based on several hundred individual underlying variables, taken from a wide variety of existing data sources. The data reflect the views on governance of survey respondents and public, private, and NGO sector experts worldwide. The WGI also explicitly report margins of error accompanying each country estimate. These reflect the inherent difficulties in measuring governance using any kind of data. Even after taking these margins of error into account, the WGI permit meaningful cross-country and over-time comparisons. The aggregate indicators, together with the disaggregated underlying source data, are available at www.govindicators.org.

This paper—a product of the Macroeconomics and Growth Team, Development Research Group—is part of a larger effort in the department to study the causes and consequences of good governance for development. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at akraay@worldbank.org.

The Worldwide Governance Indicators:

Methodology and Analytical Issues

Daniel Kaufmann, Brookings Institution

Aart Kraay and Massimo Mastruzzi, World Bank

Access the WGI data at www.govindicators.org

---

dkaufmann@brookings.edu, akraay@worldbank.org, mmastruzzi@worldbank.org. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the Brookings Institution, the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. The Worldwide Governance Indicators (WGI) are not used by the World Bank for resource allocation. Financial support from the World Bank's Knowledge for Change trust fund, and the Hewlett Foundation is gratefully acknowledged. We would like to thank S. Rose, S. Radelet, C. Logan, M. Neumann, N. Meisel, J. Ould-Auodia, R. Fullenbaum, M. Seligson, F. Marzo, C. Walker, P. Wongwan, V. Hollingsworth, S. Hatipoglu, D. Cingranelli, D. Richards, M. Lagos, R. Coutinho, S. Mannan, Z. Tabernacki, J. Auger, L. Mootz, N. Heller, G. Kisunko, J. Rodriguez Mesa, J. Riano, V. Penciakova, and D. Cieslikowsky for providing data and comments, and answering our numerous questions. Particular thanks is due to Arseny Malov for his work in designing and maintaining the WGI website at www.govindicators.org.

1. Introduction

The Worldwide Governance Indicators (WGI) are a long-standing research project to develop cross-country indicators of governance. The WGI consist of six composite indicators of broad dimensions of governance covering over 200 countries since 1996: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. These indicators are based on several hundred variables obtained from 31 different data sources, capturing governance perceptions as reported by survey respondents, non-governmental organizations, commercial business information providers, and public sector organizations worldwide.

This paper summarizes the methodology and key analytical issues relevant to the overall WGI project. The updated data for the six indicators, together with the underlying source data and the details of the 2010 update of the WGI, are not discussed in this paper but are available online at www.govindicators.org. We also plan to release and document subsequent updates of the WGI purely online, with this paper serving as a guide to the overall methodological issues relevant to the WGI project and future updates.

In the WGI we draw together data on perceptions of governance from a wide variety of sources, and organize them into six clusters corresponding to the six broad dimensions of governance listed above. For each of these clusters we then use a statistical methodology known as an Unobserved Components Model to (i) standardize the data from these very diverse sources into comparable units, (ii) construct an aggregate indicator of governance as a weighted average of the underlying source variables, and (iii) construct margins of error that reflect the unavoidable imprecision in measuring governance.

We believe this to be a useful way of organizing and summarizing the very large and disparate set of individual perceptions-based indicators of governance that have become available since the late 1990s when we began this project. Moreover, by constructing and reporting explicit margins of error for the aggregate indicators, we enable users to avoid over-interpreting small differences between countries and over time in the indicators that are unlikely to be statistically – or practically – significant.

This emphasis on explicit reporting of uncertainty about estimates of governance has been notably lacking in most other governance datasets.[1]

While the six aggregate WGI measures are a useful summary of the underlying source data, we recognize that for many purposes, the individual underlying data sources are also of interest for users of the WGI data. Many of these indicators provide highly specific and disaggregated information about particular dimensions of governance that are of great independent interest. For this reason we make the underlying source data available together with the six aggregate indicators through the WGI website.

The rest of this paper is organized as follows. In the next section we discuss the definition of governance that motivates the six broad indicators that we construct. Section 3 describes the source data on governance perceptions on which the WGI project is based. Section 4 provides details on the statistical methodology used to construct the aggregate indicators, and Section 5 offers a guide to interpreting the data. Section 6 contains a review of some of the main analytic issues in the construction and use of the WGI, and Section 7 concludes.

2. Defining Governance

Although the concept of governance is widely discussed among policymakers and scholars, there is as yet no strong consensus around a single definition of governance or institutional quality. Various authors and organizations have produced a wide array of definitions. Some are so broad that they cover almost anything, such as the definition of *"rules, enforcement mechanisms, and organizations"* offered by the World Bank's 2002 World Development Report "Building Institutions for Markets". Others more narrowly focus on public sector management issues, including the definition proposed by the World Bank in 1992 as *"the manner in which power is exercised in the management of a country's economic and social resources for development"*. In specific areas of governance such as the rule of law, there are extensive debates among scholars over "thin" versus "thick" definitions, where the former focus narrowly on whether existing rules and laws are enforced, while the latter emphasizes more the justice of the content of the laws.

[1] The only exceptions we are aware of are that (a) the Transparency International Corruption Perceptions Index began reporting margins of error in the mid-2000s, and (b) more recently the Global Integrity Index has begun reporting measures of inter-respondent disagreement on their expert assessments of integrity mechanisms.

We draw on existing notions of governance, and seek to navigate between overly broad and narrow definitions, to define governance as *"the traditions and institutions by which authority in a country is exercised. This includes (a) the process by which governments are selected, monitored and replaced; (b) the capacity of the government to effectively formulate and implement sound policies; and (c) the respect of citizens and the state for the institutions that govern economic and social interactions among them.*" We construct two measures of governance corresponding to each of these three areas, resulting in a total of six dimensions of governance:

(a) The process by which governments are selected, monitored, and replaced:

1. Voice and Accountability (VA) – capturing perceptions of the extent to which a country's citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media.

2. Political Stability and Absence of Violence/Terrorism (PV) – capturing perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including politically-motivated violence and terrorism.

(b) The capacity of the government to effectively formulate and implement sound policies:

3. Government Effectiveness (GE) – capturing perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.

4. Regulatory Quality (RQ) – capturing perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development.

(c) *The respect of citizens and the state for the institutions that govern economic and social interactions among them:*

5. Rule of Law (RL) – capturing perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.

6. Control of Corruption (CC) – capturing perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.

We believe that this definition provides a useful way of thinking about governance issues as well as a useful way of organizing the available empirical measures of governance as described below. Yet we recognize that for other purposes, other definitions of governance may of course also be relevant. In this spirit we make the source data underlying our indicators publicly available at www.govindicators.org, and encourage users with different objectives to combine the data in different ways more suited to their needs. In the next section of the paper we describe how we use our definitions to organize a large number of empirical proxies into the six categories mentioned above.

We also note that these six dimensions of governance should not be thought of as being somehow independent of one another. One might reasonably think for example that better accountability mechanisms lead to less corruption, or that a more effective government can provide a better regulatory environment, or that respect for the rule of law leads to fairer processes for selecting and replacing governments and less abuse of public office for private gain. In light of such inter-relationships, it is not very surprising that our six composite measures of governance are strongly positively correlated across countries. These inter-relationships also mean that the task of assigning individual variables measuring various aspects of governance to our six broad categories is not clear-cut. While we have taken considerable care to make these assignments reasonably in our judgment, in some cases there is also room for debate. For this reason as well, the availability of the underlying source data is a useful feature of the WGI as it allows users with other objectives, or other conceptions of governance, to organize the data in ways suited to their needs.

3. Governance Data Sources for the WGI

In the WGI project we rely exclusively on perceptions-based governance data sources. In Section 6 below we discuss in more detail the rationale for relying on this particular type of data. Our data sources include surveys of firms and households, as well as the subjective assessments of a variety of commercial business information providers, non-governmental organizations, and a number of multilateral organizations and other public-sector bodies. Table 1 identifies the full set of 31 sources used in the 2010 update of the WGI. Each of these data sources provides us with a set of empirical proxies for the six broad categories of governance that we seek to measure. For example, a cross-country household or firm survey might provide us with data on respondents' perceptions or experiences with corruption, while a NGO or commercial data provider might provide its own

assessments of corruption based on its network of respondents. As discussed in the following section, we then combine these many different measures of corruption into a composite indicator that summarizes their common component. We follow the same process for the other five broad indicators.

Complementing Table 1 in this paper, a complete description of each of these data sources, including a description of how each of the individual variables from them is assigned to one of the six broad WGI measures, is available on the Documentation tab of www.govindicators.org. Almost all of our data sources are available annually, and we align these annual observations with the years for the WGI measures. In a few cases data sources are updated only once every two or three years. In this case, we use data lagged by one or two years from these sources to construct the estimates for more recent aggregate WGI measures. Details on these issues of timing can also be found in the full descriptions of the individual data sources.

We note also that there are small changes from year to year in the set of sources on which the WGI scores are based. These too are documented online, and reflect the reality that we introduce new data sources as they become available, and if necessary on occasion drop existing sources that stop publication or undergo other significant changes that prevent us from continuing their use in the WGI. Wherever possible we make these changes consistently for all years in the historical data as well, in order to ensure maximum over-time comparability in the WGI. Users of the WGI should therefore be aware that each annual update of the WGI supersedes previous years' versions of the data for the entire time period covered by the indicators.

The WGI data sources reflect the perceptions of a very diverse group of respondents. Several are *surveys* of individuals or domestic firms with first-hand knowledge of the governance situation in the country. These include the World Economic Forum's Global Competitiveness Report, the Institute for Management Development's World Competitiveness Yearbook, the World Bank / EBRD's Business Environment and Enterprise Performance surveys, the Gallup World Poll, Latinobarometro, Afrobarometro, and the AmericasBarometer. We refer to these as "Surveys" in Table 1.

We also capture the views of country analysts at the major multilateral development agencies (the European Bank for Reconstruction and Development, the African Development Bank, the Asian Development Bank, and the World Bank), reflecting these individuals' in-depth experience working on the countries they assess. Together with some expert assessments provided by the United States

Department of State and France's Ministry of Finance, Industry and Employment, we classify these as *"Public Sector Data Providers"* in Table 1.

A number of data sources provided by various *nongovernmental organizations*, such as Reporters Without Borders, Freedom House, and the Bertelsmann Foundation, are also included. Finally, an important category of data sources for us are *commercial business information providers*, such as the Economist Intelligence Unit, Global Insight, and Political Risk Services. These last two types of data providers typically base their assessments on a global network of correspondents with extensive experience in the countries they are rating.

The data sources in Table 1 are fairly evenly divided among these four categories. Of the 31 data sources used in 2009, 5 are from commercial business information providers; surveys and NGOs contribute 9 sources each; and the remaining 8 sources are from public sector providers. An important qualification however is that the commercial business information providers typically report data for larger country samples than our other types of sources. An extreme example is the Global Insight Business Conditions and Risk Indicators, which provides information on over 200 countries in each of our six aggregate indicators. Primarily for reasons of cost, household and firm surveys typically have much smaller country coverage, although the coverage of some is still substantial. Our largest surveys, the Global Competitiveness Report survey and the Gallup World Poll each cover around 130 countries, but several regional surveys cover necessarily smaller sets of countries. Some of the expert assessments provided by NGOs and public sector organizations have quite substantial country coverage, but others, particularly regionally-focused ones have much smaller country coverage. In 2009 for example, data from commercial business information providers account for around 34 percent of the country-year data points in our underlying source data, while surveys and NGOs contribute 20 percent each, and public sector providers contribute the remaining 26 percent of data points.

As a vital complement to the aggregate WGI measures, we also make available through the WGI website the underlying data from virtually all of the individual data sources that go into our aggregate indicators. The majority of our data sources, such as Freedom House and Reporters Without Borders have always been publicly available through the publications and/or websites of their respective organizations, and we simply reproduce them here. Several of our other sources provided by commercial risk rating agencies and commercial survey organizations are only available commercially. In the interests of greater transparency, these organizations have kindly agreed to allow us to report their proprietary data in the form in which it enters our governance indicators.

The only data sources we are unable to make fully public are the World Bank's Country Policy and Institutional Assessment (CPIA), and the corresponding assessments produced by the African Development Bank and the Asian Development Bank. This reflects the disclosure policy of these organizations and not a choice on our part. We do note however that starting in 2002 the World Bank began publishing limited information on its CPIA assessments on its external website. For the years 2002-2004 the overall CPIA ratings are reported by quintile for countries eligible to borrow from the International Development Association (IDA), the concessional lending window of the World Bank. Since 2005, the individual country scores for the IDA resource allocation index, a rating that reflects the CPIA as well as other considerations, have been publicly available for IDA-eligible countries. The African Development Bank's CPIA ratings have also been publicly available by quintile since 2004, and have been fully public since 2005, while the Asian Development Bank's scores have been fully public for its concessional borrowers since 2005. Those CPIA scores made public by these multilateral development banks are also available through our website.

All the individual variables have been rescaled to run from zero to one, with higher values indicating better outcomes. These individual indicators can be used to make comparisons of countries over time, as all of our underlying sources use reasonably comparable methodologies from one year to the next. They also can be used to compare the scores of different countries on each of the individual indicators, recognizing however that these types of comparisons too are subject to margins of error. We caution users however not to compare directly the scores from different individual sources for a single country, as these are not comparable. For example, a developing country might receive a score of 0.7 on a 0-1 scale from one data source covering only developing countries, but might receive a lower score of 0.5 on the same 0-1 scale from a different data source that covers both developed and developing countries. This difference in scores could simply be due to the fact that the reference group of comparator countries is different for the two data sources, rather than reflecting any meaningful difference in the assessment of the country by the two sources. As discussed in detail in the following section, our procedure for constructing the six aggregate WGI measures provides a way of adjusted for such differences in units that allows for meaningful aggregation across sources.

4. Constructing the Aggregate WGI Measures

We combine the many individual data sources into six aggregate governance indicators, corresponding to the six dimensions of governance described above. We do this using a statistical tool known as an unobserved components model (UCM).[2] The premise underlying this statistical approach is straightforward – each of the individual data sources provides an imperfect signal of some deeper underlying notion of governance that is difficult to observe directly. This means that, as users of the individual sources, we face a signal-extraction problem – how do we isolate an informative signal about the unobserved governance component common to each individual data source, and how do we optimally combine the many data sources to get the best possible signal of governance in a country based on all the available data? The UCM provides a solution to this signal extraction problem.

For each of the six components of governance defined above, we assume that we can write the observed score of country j on indicator k, y_{jk}, as a linear function of unobserved governance in country j, g_j, and a disturbance term, ε_{jk}, as follows:

(1) $$y_{jk} = \alpha_k + \beta_k(g_j + \varepsilon_{jk})$$

where α_k and β_k are parameters which map unobserved governance in country j, g_j, into the the observed data from source k, y_{jk}. As an innocuous choice of units, we assume that g_j is a normally-distributed random variable with mean zero and variance one.[3] This means that the units of our aggregate governance indicators will also be those of a standard normal random variable, i.e. with zero mean, unit standard deviation, and ranging approximately from -2.5 to 2.5. The parameters α_k and β_k reflect the fact that different sources use different units to measure governance. For example, one data source might measure corruption perceptions on a scale from zero to three, while another might do so on a scale from one to ten. Or more subtly, two data source might both use a scale notionally running from zero to one, but the convention of one source might be to use the entire scale, while on another source scores are clustered between 0.3 and 0.7. These differences in explicit and implicit choice of units in the observed data from each source are captured by differences across sources in the parameters α_k and β_k. As discussed below, we can then use estimates of these parameters to rescale the data from each source into common units.

[2] 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)

[3] See Kaufmann, Kraay and Zoido-Lobaton (1999a) for a discussion of alternative choices of units for governance.

We assume that the error term is also normally distributed, with zero mean and a variance that is the same across countries, but differs across indicators, i.e. $V[\varepsilon_{jk}] = \sigma_k^2$. We also assume that the errors are independent across sources, i.e. $E[\varepsilon_{jk}\varepsilon_{jm}] = 0$ for source m different from source k. This identifying assumption asserts 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. In Section 6 below we discuss the likelihood and consequences of potential violations of this identifying assumption in more detail.

The error term ε_{jk} 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 could be 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 considered to be. Both of these sources of uncertainty are reflected in the indicator-specific variance of the error term, σ_k^2. The smaller is this variance, the more precise a signal of governance is provided by the corresponding data source.

Given estimates of the parameters of the model, α_k, β_k, and σ_k^2, we can now construct estimates of unobserved governance g_j, given the observed data y_{jk} for each country. In particular, the unobserved components model allows us to summarize our knowledge about unobserved governance in country j using the distribution of g_j conditional on the observed data y_{jk}. This distribution is also normal, with the following mean:

$$E[g_j|y_{j1}, \ldots, y_{jK}] = \sum_{k=1}^{K} w_k \frac{y_{jk} - \alpha_k}{\beta_k} \quad (2)$$

We use this conditional mean as our estimate of governance. It is simply a weighted average of the rescaled scores for each country, $\frac{y_{jk}-\alpha_k}{\beta_k}$. This rescaling puts the observed data from each source into the common units we have chosen for unobserved governance. The weights assigned to each source k

are given by $w_k = \frac{\sigma_k^{-2}}{1+\sum_{k=1}^{K}\sigma_k^{-2}}$, and are larger the smaller the variance of the error term of the source. In other words, sources that provide a more informative signal of governance receive higher weight.

A crucial observation however is that there is unavoidable uncertainty around this estimate of governance. This uncertainty is captured by the standard deviation of the distribution of governance conditional on the observed data:

(3) $$SD\left[g_j|y_{j1},\ldots,y_{jK}\right] = \left(1+\sum_{k=1}^{K}\sigma_k^{-2}\right)^{-1/2}$$

This standard deviation is smaller the more data sources are available for a country, i.e. the larger is K, and the more precise those individual data sources are, i.e. the smaller is σ_k^2. We refer to this number as the "standard error" of our estimate of governance for each country. These standard errors are essential to the correct interpretation of our estimates of governance, as they capture the inherent uncertainty is measuring governance. For example, whenever we compare estimates of governance for two countries, or for a single country over time, we always report the 90 percent confidence interval associated with both estimates of governance, i.e. the estimate of governance +/- 1.64 times its standard deviation. This range, which we refer to as the "margin of error" for the governance score, has the following interpretation: based on the observed data, we can be 90 percent confident that the true, but unobserved, level of governance for the countries lies in this range. A useful (and conservative) rule of thumb is that when these margins of error overlap for two countries, or for two points in time, then the estimated differences in governance are too small to be statistically significant.[4]

The presence of margins of error in our governance estimates is not a consequence of our use of subjective or perceptions-based data to measure governance. Rather, it simply reflects the reality that available data are imperfect proxies for the concepts that we are trying to measure. Just as alternative survey-based measures are imperfect proxies for the overall level of corruption in a country, fact-based description of the legal regulatory framework are also only imperfect proxies for the overall business environment facing firms. A key strength of the WGI is that we explicitly recognize this imprecision, and the margins of error we report provide users of the WGI with tools to take this imprecision into account when making comparisons between countries and over time.

[4] See Kaufmann, Kraay and Mastruzzi (2006a), Section 2.2, for details on testing for significance of over time changes in governance.

In order to construct these estimates of governance and their accompanying standard errors, we require estimates of all of the unknown survey-specific parameters, α_k, β_k, and σ_k^2 . We obtain these in a modified maximum likelihood procedure detailed in the Appendix. We estimate a new set of parameters for each year, and all of the parameter estimates for each data source in each year, together with the resulting weights, are reported online in the Documentation tab of www.govindicators.org. In addition, for each country and for each of the six aggregate indicators, we report the estimate of governance, i.e. the conditional mean in Equation (2), the accompanying standard error, i.e. the conditional standard deviation in Equation (3), and the number of data sources on which the estimate is based.

One important feature of our choice of units for governance is that we have assumed that the world average is the same in each year. While our indicators can be meaningfully used to compare countries' relative positions in a given year, and their relative positions over time, the indicators are not informative about trends in global averages of governance. While at first glance this may appear restrictive, by reviewing the time series of the individual sources over the past several updates of the WGI, we have documented that there is very little evidence of trends over time in global averages of our individual underlying data sources. As a result, our choice of units for governance which fixes the global average to be the same in each period does not appear unreasonable. Moreover, this implies that changes in countries' relative positions are unlikely to be very different from changes over time in countries' absolute positions. And finally, fixing the global average to equal zero does not prevent the analysis of trends over time in regional or other group averages of countries.

5. Using and Interpreting the WGI Data

We report the aggregate WGI measures in two ways: in the standard normal units of the governance indicator, ranging from around -2.5 to 2.5, and in percentile rank terms ranging from 0 (lowest) to 100 (highest) among all countries worldwide.[5] Figure 1 shows the data in these two ways, for two of the aggregate WGI measures in 2009, Government Effectiveness and Control of Corruption. We order countries in ascending order according to their point estimates of governance in 2009, and we

[5] Note that when we present the data in percentile rank form on the governance indicators website, we also show 90 percent confidence intervals converted into percentile rank form. In particular, the upper (lower) end of the confidence intervals in percentile rank terms is computed by calculating the percentile rank among all country scores of the upper (lower) bound of the confidence interval in standard normal units.

plot their percentile ranks on the horizontal axis, and their estimates of governance and associated 90% confidence intervals on the vertical axis.

For illustrative purposes we have labeled 20 countries equally spaced through the distribution of governance (i.e. at the 5^{th}, 10^{th}, 15^{th}, etc. percentiles of the distribution). The size of the confidence intervals varies across countries, as different countries are covered by a different numbers of sources, with different levels of precision. A key observation is that the resulting confidence intervals are substantial relative to the units in which governance is measured. 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 at reasonable confidence levels, since the associated 90 percent confidence intervals are likely to overlap.

For example, while a country such as Peru ranks ahead of a country such as Jamaica on Control of Corruption, the confidence intervals for the two countries overlap substantially, and so one should not interpret the WGI data as signaling a statistically significant difference between the two countries. For many applications, instead of merely observing the point estimates, it is often 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), recognizing that these likely ranges may overlap for countries that are being compared.

This is not to say however that the aggregate indicators cannot be used to make cross-country comparisons. To the contrary, there are a great many pair-wise country comparisons that do point to statistically significant, and likely also practically meaningful, differences across countries. For example, the 2009 Control of Corruption indicator covers 211 countries, so that it is possible to make a total of 22,155 pair-wise comparisons of corruption across countries using this measure. For 63 percent of these comparisons, 90% confidence intervals do not overlap, signaling statistically significant differences in the indicator across countries. And if we lower our statistical confidence level to 75 percent, which may be quite adequate for some applications, we find that 73 percent of all pair-wise comparisons identify statistically significant differences.

We now turn to the changes over time in our estimates of governance in individual countries. Figure 2 illustrates these changes for two selected governance indicators over the decade 2000-2009, Voice and Accountability and Rule of Law. In both panels, we plot the 2000 score on the horizontal axis, and the 2009 score on the vertical axis. We also plot the 45-degree line, so that countries located above

this line correspond to improvements in the WGI estimates of governance, while countries below the line correspond to deteriorations. The first feature of this graph is that most countries are clustered close to the 45-degree line, indicating that changes in our estimates of governance in most countries are relatively small even over the decade covered by the graph. A similar pattern emerges for the other four dimensions of governance (not shown in Figure 2), and, not surprisingly, the correlation between current and lagged estimates of governance is even higher when we consider shorter time periods than the decade shown here.

Nevertheless, a substantial number of countries do show significant changes in governance. In order to assess whether the change over time in an indicator for a given country over a certain period is significant, a useful rule of thumb is to check whether the 90 percent confidence intervals for the two periods do not overlap. We highlight and label these cases in Figure 2. More generally, over the decade 2000-2009 covered in Figure 2, we find that for each of our six indicators, on average 8 percent of countries experience changes that are significant at the 90 percent confidence level. Looking across all six indicators, 28 percent of countries experience a significant change in at least one of the six dimensions of governance over this period. Since the world averages of the WGI are constant over time, these changes are necessarily roughly equally divided between improvements and deteriorations.

We also note that the 90 percent confidence level is quite high, and for some purposes a lower confidence level, say 75 percent, would be appropriate for identifying changes in governance that may be practically important. Not surprisingly this lower confidence level identifies substantially more cases of significant changes: for the period 2000-2009, and looking across the six aggregate WGI measures, on average 18 percent of countries experience a significant change in the WGI. Over the same period, 54 percent of countries experience a significant change on at least one of the six WGI measures.

When interpreting differences between countries and over time in the six aggregate WGI measures, it is important to also consult the underlying source data. This is because differences across countries or over time in the aggregate WGI measures reflect not only differences in countries' scores on the underlying source data, but also differences in the set of underlying data sources on which the comparison is based, and in the case of changes over time, differences over time in the weights used to aggregate the indicators. In the following section we discuss in more detail the role of such compositional and weighting effects. For now, we note that to facilitate this consultation of individual indicators, access to the underlying source data is provided interactively and in downloadable format at www.govindicators.org. The underlying indicator data reported here has all been rescaled from the

original sources to run from zero to one, with higher values corresponding to better governance outcomes. Since all of our sources use reasonably comparable methodologies over time, the data from the individual indicators can usefully be compared both across countries within a given time period, and over time for individual countries. However, we caution users of the WGI data not to compare the individual indicator data from one source with another. As noted in the previous section, different indicators use different implicit as well as explicit choice of units in measuring governance. While the process of aggregation corrects for these differences, the underlying source data, even when re-scaled to run from zero to one, still reflects these differences in units and so is not comparable across sources.

6. Analytical Issues

In this section we review a number of methodological and interpretation issues in the construction and use of the WGI. Our objective is to concisely summarize a number of key points that have come up over the past decade in the WGI project, and that we have addressed in detail in our earlier papers, as referenced below. We first discuss a number of issues related to our choice of aggregation methodology. We then discuss the strengths and potential drawbacks of the subjective or perceptions-based data on which we rely to construct the WGI.

6.1 Aggregation Methodology

A first basic question one might ask is why we have chosen to use the unobserved components (UCM) methodology to construct the WGI, as opposed to other, possibly more straightforward, methods. For example a simple alternative method would be to average together the percentile ranks of countries on the individual indicators, as has been done by Transparency International in the construction of the Corruption Perceptions Index or by the Doing Business Project in the construction of the Ease of Doing Business rankings.[6] Another alternative would be to do a min-max rescaling of the source data and then average the rescaled data, as for example is done by the Ibrahim Index of African

[6] As discussed further in Appendix A, a refinement is required when the data come from sources that cover different samples of countries. For example, obtaining the top rank in a sample of developing countries may not correspond to the same level of governance performance as obtaining top rank in a sample of industrialized countries. In this case a slightly modified ranking procedure such as percentile matching is required, as for example is done by Transparency International. When all the individual sources cover the same set of countries, it is also possible to simply average the country ranks, rather than percentile ranks). This is done by Doing Business.

Governance. While the UCM methodology we use is somewhat more complex, relative to these alternatives we find it offers three main advantages.

First, the UCM's approach to placing data in common units, as described in Section 4, has the advantage of maintaining some of the cardinal information in the underlying data. In contrast, methods based on country rankings by definition retain only information on countries' relative ranks, but not on the size of the gaps between countries. And in contrast with the min-max method, the UCM approach has the advantage of being less sensitive to extreme outliers in the data.

Second, the UCM approach provides a natural framework for weighting the rescaled indicators by their relative precision, as opposed to simply constructing unweighted averages as is done by most other cross-country composite measures of governance and the investment climate. The data-driven precision-weighting approach in the UCM has the advantage of improving the precision of the overall aggregate indicators. However, we do not want to overstate the importance of this benefit. We have found that precision-weighting reduces the margins of error of our aggregate indicators by only about 20 percent relative to unweighted averages. Similarly, we find that the choice of weighting scheme has for the most part rather small effects on the ranking of countries.[7] This is largely due to the fact that the various individual indicators underlying the WGI are quite highly correlated, and so there is limited scope for changes in country rankings due to reweighting of sources.

The third advantage of the UCM methodology is that it naturally emphasizes the uncertainty associated with aggregate indicators of governance. The UCM usefully formalizes the issue of aggregation as a signal extraction problem: since "true" governance is difficult to observe and we can observe only imperfect indicators of it, how can we best extract a "signal" of unobserved governance from the observed data? Under this view, all individual indicators of corruption, for example, should be viewed as noisy or imperfect proxies for corruption. Aggregating these together can result in a more informative signal of corruption. But even these aggregate measures are imperfect and this imperfection is usefully summarized by the standard errors and confidence intervals generated by the UCM.

Moreover, by formulating the process of aggregation as a signal extraction problem, we think the UCM provides a rationale for a more inclusive approach to combining data from different types of sources. Taken at face value, a firm survey question about the prevalence of "additional payments to

[7] For details on this point refer to Kaufmann, Kraay and Mastruzzi (2007a).

get things done" and an expert assessment of public sector corruption, could be measuring quite different things. But our interpretation, and rationale for combining the two with others in a composite measure, is simply that both provide noisy or imperfect "signals" of the prevalence of corruption, and by combining information from the two, we can get a better estimate of overall corruption. Of course this comes at the cost of losing the specific nuances of the individual sources and their definitions. But this is not an "either/or" decision, as both the composite summary indicators as well as the underlying individual measures are available in the case of the WGI.

A further question related to aggregation is whether it makes sense to use all available data sources for all countries, as opposed to using only those data sources that cover all countries and in all time periods. The advantage of the latter approach is that all comparisons, both across countries and over time, would be fully "balanced" in the sense of relying on the same set of data sources. In contrast, in the WGI most comparisons over time and across countries are "unbalanced" in the sense that they are based on the potentially different sets of source data available for the two comparators. This runs the risk that changes in scores on the aggregate indicators reflect not only changes in the underlying source data available for the countries, but also changes in the set of available data sources. While relying on a purely balanced set of comparisons has the appeal of easy interpretation, this must be set against a very important practical drawback – that this restriction greatly limits the set of countries and data sources that could be covered by the aggregate indicators. For example, the 2009 Control of Corruption indicator covers 211 countries and is based on a total of 22 data sources. However, if we were to restrict attention only to a balanced sample of countries covered in the five data sources with the largest country coverage, our sample of countries would fall nearly by half to 114. Since the 22 individual data sources used by the WGI together provide 1950 country/indicator data points, restricting attention to this balanced sample would also involve discarding more than three-quarters of data sources, and nearly three-quarters of the available data points (i.e. 1-5x114/1950=0.71). Doing so would also reduce the diversity of sources, as four of the five top data sources by country coverage come just from one type, commercial business information providers.

Set against this major drawback, we think the benefits of relying on as much information as possible when constructing the aggregate indicators are worth the cost of having to interpret unbalanced comparisons over time and across countries. Moreover, it is useful to note that the extent of "unbalancedness" and its consequences are in most cases likely to be modest. For example, among the 211 countries in the 2009 Control of Corruption Indicator, there are 22,155 potential pairwise

comparisons. The median pairwise comparison would be based on 6 common data sources, and 60 percent of all pairwise comparisons would be based on at least 5 common data sources. Turning to over time comparisons, in past updates of the WGI we have documented that the large majority of statistically significant changes over time in the WGI are largely due to changes in the underlying source data, rather than to changes in the composition of data sources in the two periods.[8]

6.2 Use of Perceptions Data

As noted in the introduction, the WGI project is based exclusively on subjective or perceptions-based measures of governance, take from surveys of households and firms as well as expert assessments produced by various organizations. This decision is based on our view that perceptions data have particular value in the measurement of governance.[9] First, perceptions matter because agents base their actions on their perceptions, impression, and views. If citizens believe that the courts are inefficient or the police are corrupt, they are unlikely to avail themselves of their services. Similarly, enterprises base their investment decisions - and citizens their voting decisions - on their perceived view of the investment climate and the government's performance. Second, in many areas of governance, there are few alternatives to relying on perceptions data. For instance, this has been particularly the case for corruption, which almost by definition leaves no "paper trail" that can be captured by purely objective measures.

Third, we note that even when objective or fact-based data are available, often such data may capture the *de jure* notion of laws "on the books", which often differs substantially from the *de facto* reality that exists "on the ground". In fact, in Kaufmann, Kraay and Mastruzzi (2005) we document sharp divergences between *de jure* and *de facto* measures of business entry regulation and find that corruption is important in explaining the extent to which the former differ from the latter. Similarly, Hallward-Driemeier, Khun-Jush, and Pritchett (2010) document that there is little correspondence between firms' actual experiences with the regulatory environment and the formal *de jure* regulations that firms face in a sample of African countries. Or to take an even starker example, in every one of the 70 countries covered in the 2007 and 2008 waves of the Global Integrity Index, it is formally illegal for a

[8] See for example the discussion around Table 5 in Kaufmann, Kraay and Mastruzzi (2009), and the same discussion in previous updates of the WGI. See also the discussion around Critique 2 in Kaufmann, Kraay and Mastruzzi (2007).

[9] It is also important to keep in mind that the distinction between "subjective" and "objective" data is often less clear-cut than it might seem. "Objective" measures of governance are often based on the coding of formal laws and regulations, but even this requires judgment as the net effect of potentially conflicting rules.

public official to accept a bribe. Yet, despite them being identical when measured *de jure*, there are large differences across these countries in perceptions of the frequency with which bribes are in fact accepted by public officials.

Despite these advantages, one might nevertheless reasonably be concerned about various potential problems in the interpretation of the subjective data we rely on in the WGI. Broadly such concerns question the extent to which perceptions data adequately capture the relevant reality. A first basic issue is simply that perceptions data on governance are imprecise. This by itself is not surprising – as we have argued above, all measures of governance and the investment climate are necessarily imprecise proxies for the broader concepts they are intended to measure. Imprecision alone does not disqualify the use of perceptions-based data on governance – rather it underscores the importance of using empirical methods to quantify and take seriously the extent of imprecision, as we do with the margins of error reported in the WGI.

A potentially more serious concern is that there are various systematic biases in perceptions data on governance. One possibility is that different types of respondents differ systematically in their perceptions of the same underlying reality. For example, it could be the case that businesspeople, represented by owners of the businesses covered in a survey, or the expert assessments provided by commercial business information providers, have different views of what constitutes good governance than other types of respondents, such as households or public sector agencies. We addressed this concern directly in Kaufmann, Kraay and Mastruzzi (2007a,b), where we compared the responses of businesspeople to other types of respondents and found little in the way of significant differences in cross-country comparisons based on these two types of responses. Another possibility is that biases are introduced by the ideological orientation of the organization providing the subjective assessments of governance. We investigated this possibility in Kaufmann, Kraay and Mastruzzi (2004) by asking whether expert assessments provided by a number of rating agencies were systematically different in countries with left- or right-wing governments. Here too we found little evidence of such biases.

Another type of bias in perceptions data might be the possibility that subjective assessments of governance are driven by factors other than governance itself, such as the level of development or recent economic performance of a country. Crudely put, this concern about "halo effects" is that raters might conclude that governance in a country must be good simply because the economy is growing fast or the country is rich. While this concern seems plausible *a priori*, we show in Kaufmann, Kraay and Mastruzzi (2004, 2007b) that in practice it does not withstand empirical scrutiny.

Yet another potential source of bias comes from the possibility that different providers of governance perceptions data rely on each other's assessments, and as a result make correlated perceptions errors. This would undermine the information content in such indicators. And more subtly, it would also undermine the validity of our weighting scheme in the WGI, which is based on the observed correlations among sources. If data sources are correlated merely because they make correlated perceptions errors, it would not be appropriate to assign higher weight to such measures.[10] Assessing the practical importance of this concern is difficult because the high correlation between governance perceptions rankings from different sources could be due either to perception errors, or due to the fact that these sources are in fact accurately measuring cross-country corruption differences and so necessarily agree with each other. In Kaufmann, Kraay and Mastruzzi (2007c) we proposed a novel way to isolate these two potential sources of correlation, by comparing the ratings produced by commercial risk rating agencies (that are often thought to be most prone to such "group-think") with cross-country firm survey responses. Our striking finding was that these data sources were no more correlated among themselves than they were with the firm survey responses, casting doubt on the practical importance of this sort of bias.

7. Conclusions

In this paper we have summarized the key features of the Worldwide Governance Indicators project. The WGI project reports composite indicators of six dimensions of governance, covering over 200 countries and territories since 1996, and is updated annually. The six aggregate governance indicators are based on hundreds of individual underlying variables from dozens of different data sources. The source data underlying the WGI come from a large number of individual sources, and reflect the views on governance of thousands of survey respondents and public, private, and NGO sector experts worldwide. The underlying source data capturing this wide diversity of views and experiences is available together with the six aggregate WGI measures at www.govindicators.org.

Due to the inherently unobservable nature of the true level of governance in a country, any observed empirical measure of governance will only be an imperfect proxy for the broader dimensions of governance that it reflects, and the data informing the WGI are no exception. The practical

[10] And of course, the estimated standard errors would be larger as well since each data source would be less informative about governance. See Kaufmann, Kraay, and Zoido-Lobaton (1999a) for further discussion.

consequence of this is that our estimates of governance are subject to nontrivial margins of error. Since the beginning of this research project in the 1990s, we have emphasized the importance of the estimation, disclosure, and use of the margins of error in interpreting country scores. In particular, users of the WGI should not over-interpret small differences in performance (across countries or over time) in the six aggregate WGI measures in particular, and all governance measures in general. Careful interpretation of this data with due regard to margins of error is important, as data-based monitoring governance performance around the world has become more common, and as empirical measures of governance become more widely used by policy-makers, analysts, journalists, risk rating agencies, and multilateral and bilateral donor aid agencies

The presence of margins of error does not imply that the WGI cannot be used to make meaningful comparisons of governance across countries or over time. Rather, our estimation of, and emphasis on, such margins of error is intended to enable users to make more sophisticated use of imperfect information. Using the WGI, we find that even after taking margins of error into account, it is possible to make many meaningful cross-country and over-time comparisons: almost two-thirds of all cross-country comparisons in 2009 result in highly-significant differences (at 90 percent confidence levels), and more than one-quarter of countries show a significant change in at least one of the six WGI measures during the decade 2000-2009.

This paper has also offered a concise summary of some of the key methodological and analytical issues that come up in the construction and interpretation of composite governance indicators based on perceptions data. We refer readers to previous years' versions in the "Governance Matters" series of working papers for more detail on each of these issues. Finally, and as in past years, we continue to caution users that aggregate indicators such as the six WGI measures are often a blunt tool for policy advice at the country level. Users of the aggregate indicators can usefully complement their analysis with an in-depth examination of the the detailed disaggregated data sources underlying the WGI, together with a wealth of possible more detailed and nuanced sources of country-level data and diagnostics on governance issues.

References

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. 66:807-815.

Efron, Bradley and Carl Morris (1972). "Limiting the Risk of Bayes and Empirical Bayes Estimators – Part 1: The Empirical Bayes Case". Journal of the American Statistical Association. 67:130-39.

Goldberger, A. (1972). "Maximum Likelihood Estimation of Regressions Containing Unobservable Independent Variables". International Economic Review. 13:1-15.

Hallward-Driemeier, Mary, Gita Khun-Jush, and Lant Pritchett (2010). "Deals Versus Rules: Policy Implementation Uncertainty and Why Firms Hate It". NBER Working Paper No. 16001.

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobatón (1999a). "Aggregating Governance Indicators." World Bank Policy Research Working Paper No. 2195, Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobatón (1999b). "Governance Matters." World Bank Policy Research Working Paper No. 2196, Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobatón (2002). "Governance Matters II – Updated Indicators for 2000/01." World Bank Policy Research Working Paper No. 2772, Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2004). "Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002". World Bank Economic Review. 18:253-287.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2005). "Governance Matters IV: Governance Indicators for 1996-2004. World Bank Policy Research Working Paper No. 3630. Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2006a). "Measuring Governance Using Perceptions Data", in Susan Rose-Ackerman, ed. Handbook of Economic Corruption. Edward Elgar.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2006b). "Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2005". World Bank Policy Research Working Paper No. 4012. Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2007a). "The Worldwide Governance Indicators Project: Answering the Critics". World Bank Policy Research Working Paper No. 4149. Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2007b). "Growth and Governance: A Reply/Rejoinder". Journal of Politics. 69(2):555-562.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2007c). "Governance Matters VI: Aggregate and Individual Governance Indicators for 1996-2006". World Bank Policy Research Working Paper No. 4280. Washington, D.C.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2008). "Governance Matters VII: Aggregate and Individual Governance Indicators for 1996-2007". World Bank Policy Research Working Paper No. 4654. Washington, D.C.

Kaufmann, Daniel and Aart Kraay (2008). "Governance Indicators: Where Are We and Where Should We Be Going?" World Bank Research Observer. Spring 2008.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2009). "Governance Matters VIII: Aggregate and Individual Governance Indicators for 1996-2008". World Bank Policy Research Working Paper No. 4978. Washington, D.C.

Appendix: Estimating the Parameters of the Unobserved Components Model

In order to implement Equations (2) and (3) in the main text, we need to first estimate the unknown parameters α_k, β_k, and σ_k^2 for every indicator k. This in turn requires us to distinguish between "representative" and "non-representative" indicators, which we treat differently in the estimation process. Representative indicators are indicators that cover a set of countries in which the distribution of governance is likely to be similar to that in the world as a whole. Practically these include all of our indicators with large cross-country coverage of developed and developing indicators.

In contrast non-representative indicators cover either specific regions (for example the BEEPS survey of transition economies or the Latinobarometer survey of Latin American countries), or particular income levels (for example the World Bank CPIA ratings that cover only developing countries). Our classification of "representative" and "non-representative" indicators is given in Table 1 of this paper.

For the set of representative indicators, we use the assumption of the joint normality of g_j and ε_{jk} write down the likelihood function of the observed data. The assumption of representativeness is crucial here because it justifies our assumption of a common distribution for governance across these different sources. As useful notation, let $\alpha = \left(\alpha_1, \dots, \alpha_{K_j}\right)$, $\beta = \left(\beta_1, \dots, \beta_{K_j}\right)$, and $\sigma^2 = \left(\sigma_1^2, \dots, \sigma_{K_j}^2\right)$, and let B and Σ be diagonal matrices with α and σ^2 on the diagonal. Using this notation, the mean of the vector of observed data for each country j, y_j, is α and the variance is $\Omega = \beta\beta' + B\Sigma B'$. The contribution to the log-likelihood of country j therefore given by:

$$\ln \mathrm{L}(\alpha, \beta, \sigma^2) \propto \ln|\Omega| + \left(y_j - \alpha\right)'\Omega^{-1}\left(y_j - \alpha\right) \tag{4}$$

Summing these over all countries j and then maximizing over the unknown parameters delivers our maximum-likelihood estimates of α_k, β_k, and σ_k^2 for every representative indicator k. Identification requires that we have a minimum of three representative indicators. Note that the number of data sources available for each country varies, and so the dimension of y_j and α is $K_j \; x \; 1$, and conformably B, Σ, and Ω are $K_j \; x \; K_j$. This way we are able to compile the likelihood function even though there potentially are missing observations for each country even among the representative indicators.

We cannot apply this method to non-representative indicators. To see why, consider the maximum-likelihood estimate of α_k for some source k. Unsurprisingly this is the mean score across countries covered by indicator k. It is straightforward to see from Equation (1) that the expected value of the sample mean of scores on indicator k is $\alpha_k + \beta_k\overline{g_k}$, where $\overline{g_k}$ denotes the average level of governance in the sample of countries covered by indicator k. For representative indicators, our choice of units for governance normalizes $\overline{g_k} = 0$ and so the sample mean delivers a consistent estimate of α_k. However, for a non-representative indicator where the average level of governance is different from the world as a whole, i.e. $\overline{g_k} \neq 0$, and so the sample mean does not provide a consistent estimate of α_k.

We can nevertheless obtain consistent estimates of the unknown parameters of the non-representative indicators by using the following simple argument. If g_j were observable, we could estimate α_k, β_k, and σ_k^2 for any indicator simply by regressing the observed scores y_{jk} on g_j. Although g_j is itself not observable, we do have an estimate of g_j based on the representative indicators. In particular, let g_j^* denote this preliminary estimate of g_j based on only on the data from the representative indicators. We can decompose this conditional mean into observed governance plus its deviation from the mean, i.e. $g_j^* = g_j + u_j$. Since u_j is independent of g, we can view g_j^* as measuring g_j with classical measurement error. It is well-known that OLS estimates of β_k from a regression of y_{jk} on g_j^* will produce downward-biased estimates due to the usual attenuation bias. In particular, the probability limit of the OLS slope coefficient is $\beta_k\left(1 - V\left[u_j\right]/V\left[g_j^*\right]\right)$. Since the variance of u_j is simply the variance of the conditional mean of g_j given in Equation (3), and since $V\left[g_j^*\right]$ is observable, we can correct the OLS coefficients for this attenuation bias to arrive at consistent estimates of the parameters of the non-representative indicators. We collect all of the parameter estimates from the representative and non-representative surveys, and insert them into the expressions in Equations (2) and (3) to arrive at estimates of governance and standard errors for each country.

Finally, there are two further rescaling steps before we arrive at the final estimates that we report. We first rescale the data to set the mean of the governance estimates to zero, and their standard deviation to one. The estimates of governance obtained from the UCM theoretically have a mean of zero, and a standard deviation slightly less than one. In any particular sample, however, the mean could be slightly different. To avoid confusion in interpreting the data, we begin by setting the mean of the governance estimates for each indicator and year to zero, and the standard deviation to one. In particular, for each indicator and year, we subtract the sample mean (across countries) from each country, and divide by the sample standard deviation (across countries). We then also divide the standard errors of the governance estimates for each country by the sample standard deviation of the governance estimates.

This first rescaling is just a renormalization of the scores, and of course has no impact on countries' relative positions on the governance indicators. It is also consistent with our choice of units for governance noted above, and notably that it has a mean of zero and a standard deviation of one in each period. If there were trends in global averages of governance over time, this choice of units would not be appropriate. However, as we have documented in Kaufmann, Kraay and Mastruzzi (2004, 2005, 2006b, 2007c, 2008, and 2009), we do not find strong evidence of significant trends in world averages of governance in our underlying indicators. We therefore think this choice of units is appropriate. Moreover, absent any changes in global averages of governance, changes over time in countries' relative positions on the WGI can also be interpreted as changes in their absolute governance scores.

The second rescaling is substantively more interesting, and addresses the fact that the sample of countries covered by our governance indicators has expanded since 1996, and quite considerably for some of our indicators. If the new countries added each year were broadly representative of the worldwide distribution of governance, this too would pose no special difficulties. However, for some of our indicators, we find that countries added in later years score on average somewhat higher than

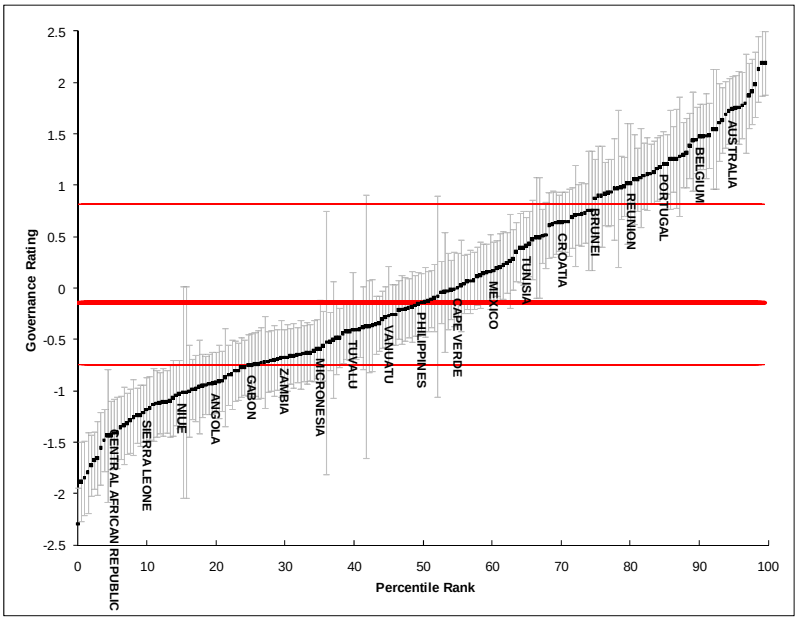
countries that were continuously in the sample. This in turn means that it would be inappropriate to impose a global average governance score of zero in earlier periods for the smaller set of countries for which data is available in the earlier periods, since our earlier estimates did not include the better-than-average performers added later. It also means that some countries in our aggregate indicators in the earlier years would have showed small declines in some dimensions of governance over time that were driven by the addition of better-performing countries in later years.

We address this issue with a second simple re-scaling of the aggregate governance indicators. We take the most recent year of our indicators as a benchmark. In particular, the indicators from 2005 onwards indicators cover between 206 and 213 countries, which are clearly representative of the world as a whole. Consistent with our choice of units for governance, the estimates for these years have zero mean and standard deviation of one across countries due to the first standardization noted above. We next consider the countries that were added in 2005 relative to 2004. We then adjust the world-wide average score in 2004 so that it would have a mean of zero *had we included the 2005 scores for those countries added in 2005 relative to 2004*.[11] We then continue backwards in time in the same way to adjust the data for each previous year back to 1996. In particular, to arrive at the final score reported for each country, we need to subtract the adjusted mean and divide by the adjusted standard deviation. The standard error of the governance estimate also needs to be divided by the indicated adjusted standard deviation. For years after 2005, we do not rescale the data in this way as the sample of countries covered by the aggregate indicators changes only minimally.

[11] The adjustment factor for the mean is simply $-\overline{y_T}(N_T - N_{T-1})/N_{T-1}$, where N_T is the number of countries with data in period T and $\overline{y_T}$ is the average score of the additional countries in period T. The higher is the average score of the new entrants and/or the more new entrants there are, the more we lower the mean in the previous period. This ensures that a hypothetical sample consisting of our year $T-1$ adjusted scores for all countries combined with the year T scores for the countries added in year T relative to $T-1$ would have a mean of zero and standard deviation of one. We also adjust the standard deviation of the year T scores to ensure that the standard deviation of this hypothetical sample would be one. We do this by multiplying the scores (and the standard errors) for each country in year $T-1$ by a factor of $\sqrt{\frac{N_T}{N_{T-1}} - \left((N_T - N_{T-1})/N_{T-1}\right)\left(V_T + \overline{y_T}^2\right) - \overline{y_{T-1}}^2}$, where V_T is the variance across countries in our estimates of governance in year T for the new entrants to the sample in period $T-1$. The greater is the dispersion in the scores of new entrants, the more we need to reduce the dispersion of scores in the previous years.

Figure 1: Governance Estimates and Margins of Error for the WGI, 2009

Government Effectiveness



Control of Corruption

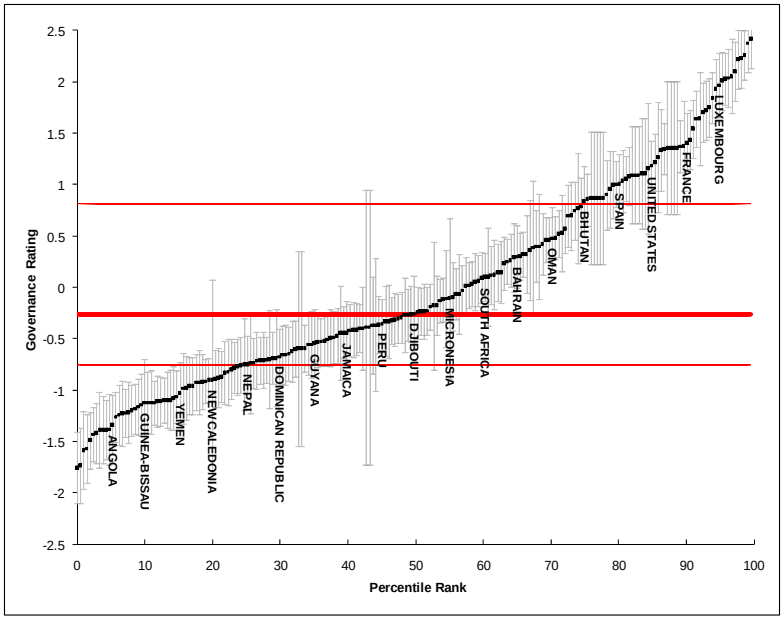
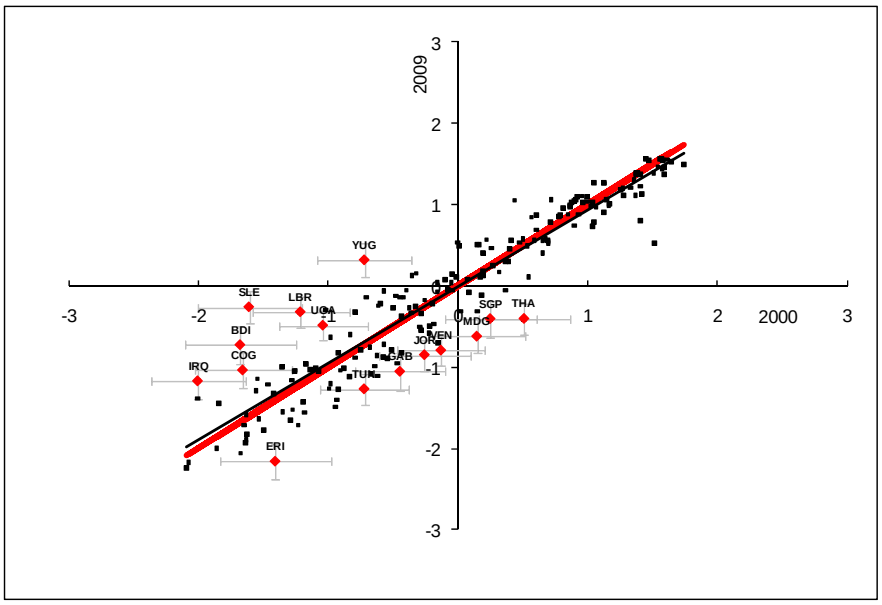


Figure 2: Changes Over Time in Governance Estimates, 2000-2009

Voice and Accountability



Rule of Law

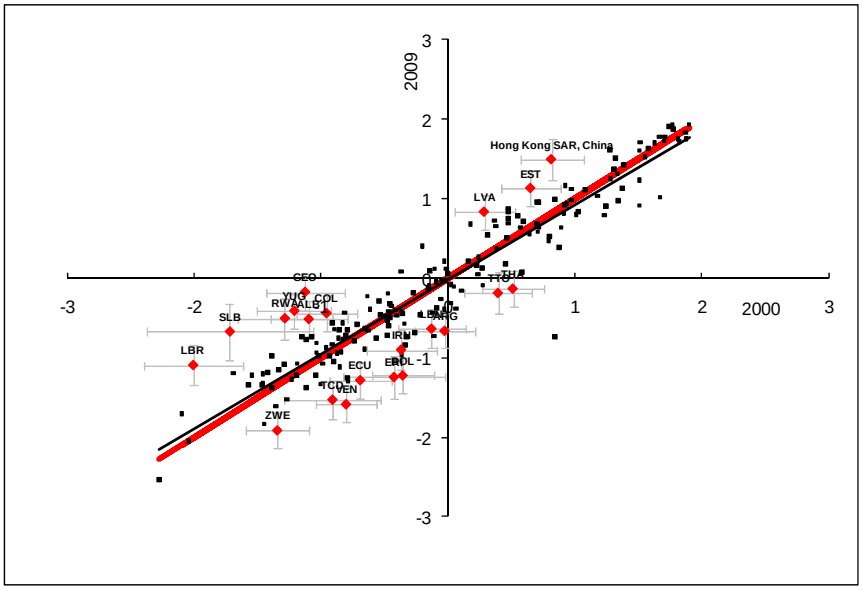


Table 1: WGI Data Sources

Code	Source	Type*	Public	Country Coverage	Represe -ntative	1996	1998	2000	2002	2003	2004	2005	2006	2007	2008	2009
ADB	African Development Bank Country Policy and Institutional Assessments	Expert (GOV)	Partial	53			x	x	x	x	x	x	x	x	x	x
AFR	Afrobarometer	Survey	Yes	19				x	x	x	x	x	x	x	x	x
ASD	Asian Development Bank Country Policy and Institutional Assessments	Expert (GOV)	Partial	29				x	x	x	x	x	x	x	x	x
BPS	Business Enterprise Environment Survey	Survey	Yes	27				x	x	x	x	x	x	x	x	x
BTI	Bertelsmann Transformation Index	Expert (NGO)	Yes	125					x	x	x	x	x	x	x	x
CCR	Freedom House Countries at the Crossroads	Expert (NGO)	Yes	62							x	x	x	x	x	x
DRI	Global Insight Global Risk Service	Expert (CBIP)	Yes	144	x	x	x	x	x	x	x	x	x	x	x	x
EBR	European Bank for Reconstruction and Development Transition Report	Expert (GOV)	Yes	29		x	x	x	x	x	x	x	x	x	x	x
EIU	Economist Intelligence Unit Riskw ire & Democracy Index	Expert (CBIP)	Yes	181	x	x	x	x	x	x	x	x	x	x	x	x
FRH	Freedom House	Expert (NGO)	Yes	197	x	x	x	x	x	x	x	x	x	x	x	x
GCB	Transparency International Global Corruption Barometer Survey	Survey	Yes	80					x	x	x	x	x	x	x	x
GCS	World Economic Forum Global Competitiveness Report	Survey	Yes	134	x	x	x	x	x	x	x	x	x	x	x	x
GII	Global Integrity Index	Expert (NGO)	Yes	79						x	x	x	x	x	x	x
GWP	Gallup World Poll	Survey	Yes	130	x								x	x	x	x
HER	Heritage Foundation Index of Economic Freedom	Expert (NGO)	Yes	179	x	x	x	x	x	x	x	x	x	x	x	x
HUM	Cingranelli Richards Human Rights Database and Political Terror Scale	Expert (GOV)	Yes	192	x	x	x	x	x	x	x	x	x	x	x	x
IFD	IFAD Rural Sector Performance Assessments	Expert (GOV)	Yes	90							x	x	x	x	x	x
IJT	iJET Country Security Risk Ratings	Expert (CBIP)	Yes	185	x						x	x	x	x	x	x
IPD	Institutional Profiles Database	Expert (GOV)	Yes	85	x								x	x	x	x
IRP	IREEP African Electoral Index	Expert (NGO)	Yes	53				x	x	x	x	x	x	x	x	x
LBO	Latinobarometro	Survey	Yes	18		x	x	x	x	x	x	x	x	x	x	x
MSI	International Research and Exchanges Board Media Sustainability Index	Expert (NGO)	Yes	76					x	x	x	x	x	x	x	x
OBI	International Budget Project Open Budget Index	Expert (NGO)	Yes	85								x	x	x	x	x
PIA	World Bank Country Policy and Institutional Assessments	Expert (GOV)	Partial	142		x	x	x	x	x	x	x	x	x	x	x
PRC	Political Economic Risk Consultancy Corruption in Asia Survey	Survey	Yes	15			x	x	x	x	x	x	x	x	x	x
PRS	Political Risk Services International Country Risk Guide	Expert (CBIP)	Yes	140	x	x	x	x	x	x	x	x	x	x	x	x
RSF	Reporters Without Borders Press Freedom Index	Expert (NGO)	Yes	170	x				x	x	x	x	x	x	x	x
TPR	US State Department Trafficking in People report	Expert (GOV)	Yes	153	x			x	x	x	x	x	x	x	x	x
VAB	Vanderbilt University Americas Barometer	Survey	Yes	23							x	x	x	x	x	x
WCY	Institute for Management and Development World Competitiveness Yearbook	Survey	Yes	55		x	x	x	x	x	x	x	x	x	x	x
WMO	Global Insight Business Conditions and Risk Indicators	Expert (CBIP)	Yes	203	x		x	x	x	x	x	x	x	x	x	x

*Types of Expert Assessments: CBIP -- Commercial Business Information Provider, GOV -- Public Sector Data Provider, NGO -- Nongovernmental Organization Data Provider