

Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2005

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi
The World Bank

September 2006

Abstract: This paper reports on the latest version of the worldwide governance indicators, covering 213 countries and territories and measuring six dimensions of governance since 1996 until end-2005: voice and accountability political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. The latest indicators are based on hundreds of variables and reflect the views of thousands of citizen and firm survey respondents and experts worldwide. Although global averages of governance display no marked trends over the period 1996-2005, nearly one-third of countries exhibit significant changes -- for better or for worse -- on at least one dimension of governance. Three new features distinguish this update. (1) We have moved to reporting estimates of governance on an annual basis. In this update we provide new estimates of governance for 2003 and 2005, as well as minor backward revisions to our bi-annual historical data for 1996-2004. (2) We are for the first time publishing the individual measures of governance from virtually every data source underlying our aggregate governance indicators. The ready availability of the individual data sources underlying the aggregate governance indicators is aimed at further enhancing the transparency of our methodology and of the resulting aggregate indicators, as well as helping data users and policy-makers identify specific governance challenges in individual countries. (3) We present new evidence on the reliability of expert assessments of governance which, alongside survey responses, form part of our aggregate measures of governance.

The aggregate and underlying individual governance indicators, a new non-technical booklet describing the data and analysis, and a number of Web-based tools for analysis of the data, are available at www.govindicators.org.

1818 H Street NW, Washington, D.C. 20433. dkaufmann@worldbank.org, akraay@worldbank.org, mmastruzzi@worldbank.org. The Governance Matters series, as well as its Worldwide Governance Indicators, are the output of a research project on governance indicators initiated in the late 1990's. The opinions expressed here, as well as the worldwide governance 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 Claudio Raddatz for helpful discussions, and A. Karatnycky, R. Fullenbaum, S. Sarkis, Jim Langston, S. Hatipoglu, D. Cingranelli, D. Richards, R. Writer, M. Gibney, C. MacCormac, E. Zukov, X. Ye, A. Gelb, G. Turner, M. Walsh, D. West, T. Sysler, B. Broadfoot, M. Lagos, A. Lopes-Claros, R. Coutinho, S. Mannan, and D. Cieslikowsky for providing data and answering our numerous questions. The support and collaboration of the World Economic Forum, the U.S. State Department, and the Netherlands Government is appreciated.

1. Introduction

This paper presents the latest update of our aggregate governance indicators.¹ The indicators measure six dimensions of governance: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. They cover 213 countries and territories for 1996, 1998, 2000, and annually for 2002-2005. The indicators are based on several hundred individual variables measuring perceptions of governance, drawn from 31 separate data sources constructed by 25 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 period. We present the point estimates of the dimensions of governance as well as the margins of errors for each country and period.

We begin by describing the data used to construct this round of the governance indicators in Section 2. As discussed in more detail below, we have added one new data source in this round. In the interests of greater comparability over time, we have included it in our past estimates as well, and also dropped a few minor data sources we have used in the past but are no longer regularly available. We have also made minor changes to the scaling of our indicators in earlier years in order to make over-time comparisons of the aggregate indicators more robust to year-to-year changes in the composition of the sample of countries included in the indicators. These revisions have resulted in minor changes to our historical biannual data for 1996-2004, and so the new dataset described here supersedes previous releases. Going forward we plan to continue to update our governance indicators on an annual basis, in order to enable a more timely monitoring of governance worldwide. While we recognize that in many countries year-over-year changes in governance are small, by moving to annual data we aim to assist users to spot and monitor those cases where there are substantial changes in the short term.

¹ This paper is the fifth in a series of estimates of governance across countries. Each of these papers has reported on data updates as well as new empirical results using the governance indicators Documentation of the previous rounds of the indicators and analytical findings can be found in Kaufmann, Kraay, and Zoido-Lobaton (1999a,b, 2002), and Kaufmann, Kraay, and Mastruzzi (2004, 2005, 2006). Kaufmann, Kraay, and Mastruzzi (2004) contains a complete description of the statistical methodology underlying the indicators.

An important innovation with this round of the governance indicators is that we are now able to fully report data from virtually all of our underlying individual sources (on the web), alongside our aggregate indicators, as we have obtained permission from the institutions generating the data to do so. This new feature will be valuable to users interested in unbundling our aggregate indicators for particular countries in order to better understand the factors contributing to our estimates of levels and changes over time in governance. Moreover, even within each of our aggregate governance indicators, the underlying measures provide a wealth of information on different specific sub-dimensions of governance. By making the underlying data readily available we hope to help users identify -- and act upon -- specific governance challenges identified by these indicators in individual countries.

As in the past, we complement our estimates of governance for each country with estimates of margins of error that indicate the unavoidable uncertainty associated with measuring governance across countries. These margins of error have declined over time with the addition of new data sources to our aggregate indicators, and are substantially smaller than for any of the individual data sources. We continue to encourage users of the governance indicators to take these margins of error into account when making comparisons of governance across countries, and within countries over time. We also encourage other producers of governance indicators to be similarly transparent about the imprecision of all types of measures of governance. Our aggregate indicators are sufficiently informative that many cross-country comparisons of governance can result in statistically -- and practically -- significant differences. In fact, using our 2005 indicators, we show that around 60 percent of all cross-country comparisons reveal strongly significant differences, and we also find that nearly one-third of the countries in our sample experience a significant change in at least one dimension of governance between 1996 and 2005. However, in other cases users will find that confidence intervals for governance based on our reported margins of error will overlap, indicating that comparisons are neither statistically -- nor practically -- significant.

We emphasize, however, that the margins of error we report are not unique to our aggregate indicators, nor are they unique to perceptions-based measures of governance on which we rely: measurement error is pervasive among all indicators of

governance and institutional quality, including individual indicators as well as so-called 'objective' or fact-based ones -- if these are available at all. Unfortunately, typically little if any effort is placed in estimating, let alone reporting, the substantial margins of error in any other source of governance and/or investment climate indicators – objective or subjective, aggregate or individual. A key advantage of our measures of governance is that we are explicit about the accompanying margins of error, whereas in most other cases they are at best left implicit, and often ignored entirely.

In the third section of this paper we consider in detail a particular critique of expert assessments of governance (which account for 23 of our 31 data sources). Some of these expert assessments are produced by commercial risk rating agencies, others by governments and multilateral organizations, and yet others by non-governmental organizations. We analyze whether each of these data sources provides independent estimates of governance, and if not, the extent to which one set of expert assessments is influenced by the views of other experts. Here we make three points. First, finding evidence of such correlated perception errors is intrinsically difficult. A high correlation in the rankings of two expert assessments could in principle be due either to highly correlated perceptions errors, or alternatively it could be due to the fact that both data sources are actually measuring cross-country differences in governance well. Second, under certain reasonable assumptions that we detail below, we can solve this identification problem and separate out the extent to which expert assessments make correlated errors. When we do this, we find at best very weak evidence of correlated perception errors, suggesting that this critique of expert assessments has little empirical basis. Third, we note that even if expert assessments do contain correlated perception errors, this does not imply that we should discard such data sources entirely. Even with correlated errors, expert assessments do contain information about cross-country differences in governance and this information can usefully be added to our aggregate governance indicators.

We conclude the paper by summarizing the key findings in this fifth installment of the Governance Matters series and noting the policy implications of our work.

2. Description of the 2005 Data Update

In this section we briefly describe the update of our governance indicators for 2005, as well as some minor backwards revisions to the biannual indicators for 1996-2004. This year we report data for 2003, in order to create a panel of annual observations covering 2002-2005. We also for the first time will be distributing on the web the bulk of data from the individual indicators of governance that underlie our aggregate indicators. We therefore also describe how this data from the individual indicators can be used to aid in the interpretation of the aggregate indicators.

Our methodology has not changed from past years, and a detailed discussion can be found in Kaufmann, Kraay, and Mastruzzi (2004). As before we construct indicators of six dimensions governance:

1. *Voice and accountability (VA)*, 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 free media
2. *Political stability and absence of violence (PV)*, perceptions of the likelihood that the government will be destabilized or overthrown by unconstitutional or violent means, including political violence and terrorism
3. *Government effectiveness (GE)*, 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)*, the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development
5. *Rule of law (RL)*, the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence
6. *Control of corruption (CC)*, 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.

In brief our methodology consists of identifying many individual sources of data on perceptions of governance that we can assign to these six broad categories. We then use a statistical methodology known as an unobserved components model to construct aggregate indicators from these individual measures. These aggregate indicators are weighted averages of the underlying data, with weights reflecting the precision of the individual data sources. Crucially our methodology also generates margins of error for the estimates of governance for each country, which need to be taken into account when making comparisons of governance across countries and over time. We provide details on this approach in the remainder of this section.

2.1 Data and Methodology

As in past years we rely on a large number of individual data sources that provide us with information on perceptions of governance. These data sources consist of surveys of firms and individuals, as well as the assessments of commercial risk rating agencies, non-governmental organizations, and a number of multilateral aid agencies. A full list of these sources is presented in Table 1. For the 2005 round of the data, we rely on a total of 276 individual variables measuring different dimensions of governance. These are taken from 31 different sources, produced by 25 different organizations. Appendices A and B provide a detailed description of each data source, and document how we have assigned individual questions from these data sources to our six aggregate indicators. Almost all of our data sources are available annually, and we use the data only the most recent year in our aggregate indicators. In a few cases, as noted in Appendix B, we use data lagged one or two years if current data are not available.²

In this round of the governance indicators we have added one new data source, the United States State Department's Trafficking in People Report. It provides country narratives and numerical rankings of countries' policy efforts to prevent trafficking in

² We never use lagged data when current data are available, in order to ensure that our indicators are as timely as possible. In this respect our approach differs from Transparency International's Corruption Perceptions Index, which in each year uses data from the current and two previous years for some of its sources (GCS, WCY and PRC), although peculiarly it does not follow the same practice for three other annually-available sources (FHT, MIG, and WMO). As a result the 2005 Corruption Perceptions Index is based on 2005 data from only 10 data sources. In contrast our Control of Corruption Index is based on 19 data sources, all of which refer to 2005.

people. Countries are ranked into three tiers, beginning in 2001. We use this tier classification as an ingredient for our Rule of Law indicator, which contains a number of other measures of the prevalence of criminal activity. In addition, a few of the data sources that we have used in the past have either been discontinued, or have not been updated and do not appear likely to be updated in the future on a regular basis. These include the Opacity Factor produced by Price-Waterhouse-Coopers, the USAID/Vanderbilt University Democracy Surveys, the Fundar rankings of budget transparency in Latin America, and the African governance indicators produced by the UN Economic Commission for Africa.³ To improve the comparability of our estimates of governance over time by having a somewhat more balanced set of sources in each period, we have dropped these measures and recalculated our historical estimates based on this slightly smaller set of data sources. Finally, we have made a number of minor revisions and corrections to the past underlying indicators for 1996-2004. Together these revisions in virtually all cases result in only minor changes in our earlier estimates of governance.⁴

Our 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's business environment surveys, and a

³ In the 2004 update of the governance indicators we relied on data from 37 sources produced by 31 organizations. Our number of organizations falls by six of our deletion of four sources, one of which was produced by two organizations (the Latin America surveys produced by USAID and Vanderbilt University), and also because Global Insight which produced the DRI ratings in 2004 has since also acquired World Markets Online, another of our sources. The number of sources falls by six because (i) we have dropped four sources, (ii) we now count three separate Gallup surveys performed in different years as one source, (iii) we no longer separately count the GCS survey of Africa in 1998 as a separate source but merge it with the GCS for all other countries in that year, and (iv) we add one new source as described in the text, for a net change of six.

⁴ The correlation between our revised and updated estimates is greater than 0.99 for all but four of the series affected by our backward revisions. Our revised data for GE in 1996 is correlated with the old data at 0.98. In 2004 our main revision was to use the final 2004 CPIA scores, which were unavailable at the time of the release of the 2004 indicators. This had a slightly greater effect on our estimates for PV, RQ, and CC, whose correlations with the old indicators range from 0.96 to 0.97. In only one small territory did our revisions result in a change in an estimate of governance that was statistically significant (in the sense of the 90% confidence intervals for the two estimates not overlapping). This is for the small island state of Tuvalu (which has very few sources) for CC in 2004, where our revisions resulted in a significant improvement for that country.

variety of global polls of individuals conducted by Gallup, Latinobarometro, and Afrobarometro. We also capture the perceptions 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. Other data sources from NGOs (such as Amnesty International, Reporters Without Borders, and Freedom House), as well as commercial risk rating agencies (such as the Economist Intelligence Unit and DRI-McGraw Hill) base their assessments on a global network of correspondents typically living in the country they are rating.

As in our past work, we combine the many individual data sources into six aggregate governance indicators. The premise underlying this statistical approach should not be too controversial – each of the individual data sources we have provides an imperfect signal of some deep 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 the informative signal about governance from 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 statistical procedure we use to perform this aggregation, known as the unobserved components model, is described in detail in our past work (see for example Kaufmann, Kraay and Mastruzzi (2004)). The main advantage of this approach is that the aggregate indicators are more informative about unobserved governance than any individual data source. Moreover, the methodology allows us to be explicit about the precision – or imprecision – of our estimates of governance in each country. As we discuss in more detail throughout the paper, this imprecision is not a consequence of our reliance on subjective or perceptions data on governance – rather imprecision is an issue that should be squarely addressed in all efforts to measure the quality of governance.

An important innovation this year is that we have obtained permission from the majority of our data sources to report the underlying indicators that go into our aggregate indicators. The sources we have made available on our website are noted in Table 1. A number of our data sources, such as Freedom House and the Heritage Foundation have always been publicly available through the publications and/or websites of their

respective organizations. Several of our other sources provided by commercial risk rating agencies and commercial survey organizations have only been available for a fee. 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. As discussed in detail in Appendix A and B, we in some cases use a simple average of multiple questions from the same source as an ingredient in our governance indicators. On our website we report either the individual question, or the average of individual questions, from each source that enters into our governance indicators. All the individual variables have been rescaled to run from zero to one, with higher values indicating better outcomes.

The only data sources we have not been able to obtain permission to publicize fully are the World Bank's Country Policy and Institutional Assessment, and the corresponding internal assessments produced by the African Development Bank and the Asian Development Bank. We do note however that starting in 2002 the World Bank has begun 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 the low-income countries eligible to borrow from the International Development Association (IDA), the soft-loan window of the World Bank. For the data covering 2005, the individual country scores for the IDA allocation factor, a rating that reflects the CPIA as well as other considerations, has now become publicly available. The African Development Bank's CPIA ratings are also publicly available by quintile only since 2004, and the Asian Development Bank has committed to publicly disclosing its ratings starting in 2005.⁵

Finally, we have this year slightly altered our presentation of the aggregate indicators in order to make them more comparable over time. In our aggregation procedure, we assume that the distribution of governance across countries follows a normal distribution in each period, with mean zero and standard deviation of one. This is simply a convenient choice of units for governance, and would be entirely innocuous were it not for two considerations. First, as we have discussed in our previous work, it is

⁵ For the African Development Bank see http://www.afdb.org/pls/portal/docs/PAGE/ADB_ADMIN_PG/DOCUMENTS/OPERATIONSINFORMATION/AFDB_2004_RATINGS_FINAL.PDF, and for the Asian Development Bank see <http://www.adb.org/Documents/Reports/ADF/2005-adb-cpa.pdf>.

possible that there are trends in world-wide averages of governance, either improvements or declines. If this is the case, it would be inappropriate to rescale the mean of governance to zero in each period as we do. However, as we have documented in the past, and also later in this paper, we do not find much evidence from our underlying individual data sources that there are trends in global averages in governance. As a result, we think the assumption of a constant global average of governance in our aggregate indicators is reasonable.

The second consideration is that the sample of countries covered by our governance indicators has expanded since 1996, and quite considerably for some of our indicators (see Table 2 and the accompanying discussion in the next subsection). 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, 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 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 simple re-scaling of the aggregate governance indicators. We take our 2005 indicators which, depending on the governance component, cover between 203 and 213 countries as representative of the world as a whole, as a benchmark. Consistent with our choice of units for governance, the estimates for 2005 have zero mean and standard deviation of one across countries. 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*. As a specific example consider Political Stability and Absence of Violence, where between 2004 and 2005 we have added 6 countries, mostly small islands in the Caribbean. Their average score in 2005 is 1.07 which is considerably above the world average of zero. We therefore slightly lower the scores of all of the 207 countries in 2004 by a factor of 0.03 in

order to improve the comparability of their scores with 2005.⁶ We then continue backwards in time in the same way to adjust the 2003 data. In particular, we look at the entrants to the sample in 2003 relative to 2005, and compute the average score for these countries, using either the 2004 estimates if they are available, or else the 2005 estimates if the country shows up in the sample only in 2005. We then again adjust the 2003 estimates so that a hypothetical sample consisting of the 2003 adjusted scores as well as either the 2004 or 2005 scores of the entrants, whichever is closer to 2003, would have a mean of zero. We continue in this way back to 1996.

Four points are worth noting about this new adjustment.

- Since we adjust the scores for all countries in a given year and indicator by the same amount, this adjustment has no effect on the relative positions of countries on that indicator in that year. It does however make countries' scores more comparable over time, since the adjustment is designed to remove the effect of adding new countries on the scores of countries already in the sample.
- As a consequence of this adjustment, global averages of the adjusted data show moderate trends over time, mostly improvements. The most extreme case is that of Political Stability and Absence of Violence where the average across all countries in the sample improves from -0.15 in 1996 to 0 in 2005. It is important to remember though that this improvement does not reflect an average

⁶ The adjustment factor for the mean is simply $-\bar{y}_T \cdot (N_T - N_{T-1})/N_{T-1}$ where N_T is the number of countries with data in period T and \bar{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 2004 by a factor of

$\sqrt{N_T/N_{T-1} - ((N_T - N_{T-1})/N_{T-1}) \left(V_T + \bar{y}_T^2 \right) - \bar{y}_{T-1}^2}$, 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. 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. The adjustments to the standard deviation are in all cases quite small, ranging from 0.96 to 1.08 with a mean adjustment factor of 1 (i.e. no adjustment at all). The adjustments to the means are somewhat higher, particularly in earlier periods where we had smaller samples, and range from -0.20 to 0.01. The interested reader can retrieve all of these adjustment factors simply by calculating the mean and standard deviation of our governance indicators in each period.

improvement for all countries in the world. Rather it reflects the changing composition of our sample since the new entrants during this period have had above-average performance in this dimension of governance. In fact, there is no evidence of any significant improvement in the world average for the country sample that has been consistently covered over time, as we discuss in more detail below.

- This rescaling of the aggregate indicators is perfectly consistent with the unobserved components model that we use to construct the aggregate indicators in each period. In particular, rescaling the mean and standard deviation of the aggregate indicators in the way that we do is equivalent to imposing slightly different means and standard deviations of governance as a choice of units in each of the periods. And as we have argued this changing choice of units is an appropriate way to correct for changes in the composition of countries covered by the indicators over time.
- Finally, for some purposes it is useful to look just at countries' percentile ranks rather than their scores on our governance indicators. Without similar adjustments these percentile ranks too would not be fully comparable over time as they too would be influenced by new entrants. Thus, we also perform such adjustment to the percentile ranks, and when we report countries' scores in the form of percentile ranks on our website, we compute the percentile ranks based on a sample consisting of the actual data we have for that indicator and year, combined with imputed data from the nearest year as described above.

2.2 Estimates of Governance 1996-2005

In Appendix C we report the aggregate governance indicators, for all countries, for each of the six indicators. The aggregate indicators, as well as almost all of the underlying indicators, are available at www.govindicators.org. The units in which governance is measured follow a normal distribution 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.⁷ This also implies that our aggregate estimates convey no information about trends in global averages of

⁷ For a handful of cases, individual country ratings can exceed these boundaries when estimates of governance are particularly high or low.

governance, but they are of course informative about changes in individual countries' relative positions over time. Below we discuss the information conveyed by some of our individual indicators regarding 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 seven periods. In 2005 the Political Stability and Absence of Violence indicator covers the largest set of 213 countries, with the other sources covering between 203 and 210 countries.⁸ Over time, there has been a steady increase in the number of sources included 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 seven to eleven in 2005. Thanks to the increase in sources, 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 7 percent of the sample in 2005.

An important consequence of this expanding 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 standard error⁹ ranged from 0.28 to 0.40. In 2005 the standard error ranges from 0.17 to 0.21 for five of our six indicators, while for Political Stability it is 0.28. These declines in margins of error illustrate the benefits in terms of precision of constructing composite indicators based on an expanding number of data sources incorporating as much information as possible. Of course, since our aggregate indicators combine information from all of these sources, they have greater precision than any individual underlying data source. Looking across all seven time periods, the median standard

⁸ A few of the entities covered by our indicators are not fully 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 all 215 entities are often referred in this paper as "countries".

⁹ As described in detail in Kaufmann, Kraay and Mastruzzi (2004), the output of our aggregation procedure is a distribution of possible values of governance for a country, conditional on the observed data for that country. The mean of this conditional distribution is our estimate of governance, and we refer to the standard deviation of this conditional distribution as the "standard error" of the governance estimate.

error of the individual data sources for the governance indicators was substantially higher at 0.58, with a 25th percentile of 0.44 and a 75th percentile of 0.85.¹⁰

Despite this increase in precision as a benefit of aggregation relative to individual data sources, and as emphasized in our previous papers, the margins of error for the aggregate governance indicators remain non-trivial. We illustrate this point in Figure 1. In the two panels of Figure 1, we organize countries in ascending order according to their point estimates of governance in 2005 on the horizontal axis, and on the vertical axis we plot the estimate of governance and the associated 90% confidence intervals. These intervals indicate the range in which it is 90 percent likely that the true governance score falls.¹¹ We do this for two of the six governance indicators, political stability, and control of corruption. 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 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 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), recognizing that these likely ranges often overlap for countries that are being compared with each other.¹²

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 pairwise country

¹⁰ In an earlier paper (Kaufmann, Kraay and Mastruzzi (2004)) we showed how to obtain margins of errors for other objective measures of governance and found that they were as large, or larger than those of our individual subjective measures. This underscores the fact that all efforts to measure governance involve margins of error, often non-trivial.

¹¹ A x% confidence interval for governance can be obtained as the point estimate of governance plus or minus the standard error times the $(100-x)/2^{\text{th}}$ percentile of the standard normal distribution. For example, the 90% confidence intervals we report throughout the paper are the point estimate plus or minus 1.64 times the standard error.

¹² Of course, asking whether 90% confidence intervals overlap or not corresponds to a hypothesis test at a significance level that is more stringent than 10%. The assumptions underlying our statistical model imply that the standard error of the difference between two country scores is the square root of the sum of the squared standard errors of the two sources, which is always smaller than the sum of the two standard errors themselves. It is more convenient -- and more conservative -- for users to simply inspect confidence intervals and see whether they overlap.

comparisons that do point to statistically significant, and likely also practically meaningful, differences across countries. Our 2005 Control of Corruption indicator for example covers 204 countries, so that it is possible to make 20,706 pairwise comparisons of corruption across countries using this measure. For 64 percent of these comparisons, 90% confidence intervals do not overlap, signaling quite highly statistically significant differences across countries. And if we lower our confidence level to 75 percent, which may be quite adequate for many applications, we find that 74 percent of all pairwise comparisons are statistically significant. The benefit of improved precision of aggregate indicators with increased data availability over time can also be clearly seen from this calculation. Consider our 1996 Control of Corruption indicator, which was based on a median of only four data sources per country, as opposed to a median of eight sources in 2005, implying substantially higher margins of error in 1996. Of the 11,476 possible pairwise comparisons in 1996, only 45 percent are significant at the 90% confidence level, and only 58 percent at the 75 percent confidence interval.

We can also use this calculation to illustrate the benefits of making comparisons based on aggregate indicators that are more informative than individual indicators. Again for Control of Corruption in 2005, consider one of our individual data sources, DRI, which has an estimated standard error of 0.52, corresponding to the median of the standard errors of all of our sources for corruption in 2005. Note that this is of course much higher than the standard error of the typical country in the aggregate indicator in 2005, which is 0.19. Had we based our estimates of governance on just this one data source, only 39 percent of cross-country comparisons would have been significant at the 75 percent level, and only 20 percent at the 90 percent level. Although rarely acknowledged explicitly, all other measures of governance are subject to margins of error as well, which in our past work we have shown to be at least as large as those we calculate for our individual and aggregate indicators. This underscores the need for caution in making cross-country comparisons with any type of governance indicator.

2.3 Changes over Time in Governance at the Country Level

We now turn to the changes over time in our estimates of governance in individual countries. In Kaufmann, Kraay, and Mastruzzi (2005) we provided a detailed analysis of how to perform statistical inference on changes over time in the aggregate

governance indicators. Here we simply provide a brief description of changes over time based on the latest update of the indicators. Figure 2 illustrates these changes for two selected governance indicators over the period 2002-2005, Political Stability and Absence of Violence/Terrorism, and Control of Corruption. In both panels, we plot the 2002 score on the horizontal axis, and the 2005 score on the vertical axis. We also plot the 45-degree line, so that countries above this line correspond to improvements in governance, while countries below the line correspond to deteriorations in governance. The first feature of this graph is that most countries are clustered quite close to the 45-degree line, indicating that changes in our estimates of governance in most countries are relatively small over the three-year period 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 three-year period shown here.

In Figure 2 we have labeled those countries for which the change in estimated governance over the 2002-2005 period is sufficiently large that the 90% confidence intervals for governance in the two periods do not overlap.¹³ Examples of such more substantial changes in governance between 2002 and 2005 include significant declines in Political Stability and Absence of Violence/Terrorism in Thailand, Bolivia and Bangladesh, and improvements in Georgia. For Control of Corruption countries such as Turkey and Georgia see significant improvements. Countries such as Kenya, Liberia, Ukraine, and Iraq all improve substantially on Voice and Accountability, while Singapore and Russia register declines.

In Table 3 we provide more detail on all of the large changes in our six governance indicators over the period 2002-2005. The first three columns report the level of governance in the two periods, and the change. The next three columns report on how the underlying data sources move for each case. 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”

¹³ While this is not a formal test of the statistical significance of changes over time in governance, it is a very simple and transparent rule of thumb for identifying large changes in governance. In a Kaufmann, Kraay, and Mastruzzi (2005, 2006) we have shown in more detail how to assess the statistical significance of changes in governance, and that this simple rule of thumb turns out to be a fairly good approximation.

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").

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 0.8 for the period 2002-2005, as reported in Table 3. This provides some confidence that for countries with large changes in our governance estimates, these changes are most often being driven primarily by changes in underlying sources rather than by the addition or deletion of sources. In fact, for these large changes there are only two cases where the agreement ratio is 50 percent or less. These cases are small Caribbean island economies that had only two data sources in 2002, and the addition of a new data source in 2005 resulted in a substantial change in these countries rankings. As can be seen from Table 3, such cases where the addition of a data source for a country significantly affects a country's score are quite rare.¹⁴

It is also worth noting that the agreement ratios for large changes in governance are substantially higher than the agreement ratios for all changes in governance. This can be seen in Table 4 which computes the same agreement ratio, but for all countries over the period 2002-2005. The agreement ratio averages 62 percent, compared with 80 percent for large changes, suggesting that for the more typical smaller changes in our governance estimates, there is relatively more disagreement across individual sources about the direction of the change than there is for large changes. These examples underscore the importance of carefully examining the factors underlying changes in the aggregate governance indicators in particular countries. In order to facilitate this, on our website users can now retrieve the data from the individual indicators underlying our aggregate indicators and use this to examine trends in the underlying data as well as changes over time in the composition of data sources on which the estimates are based.

¹⁴ The only other cases for which this occurs are the Solomon Islands and Vanuatu, where the addition of these countries to our data source WMO resulted in a large changes in these countries scores (as can also be seen in Figure 2)

While the number of countries experiencing highly significant changes in governance over the relatively short period between 2002 and 2005 is small, we do note that over longer periods, a much greater number of countries experience significant changes in governance. Our aggregate indicators now span a 10 year period from 1996-2005. Over the course of this decade, we find that on average, about 8 percent of countries experience changes that are significant at the 90 percent confidence level on each of the six indicators. Looking across all six indicators, 31 percent of countries experience a significant change at 90 percent confidence level in at least one of the six dimensions of governance over this period. 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 are likely to be practically important. Not surprisingly this lower confidence level identifies substantially more cases of significant changes: 19 percent of countries experience a significant change on each indicator on average, and fully 60 percent of countries experience a significant change on at least one dimension of governance.

As we have noted above, in some cases the addition of sources over time does have a significant impact on the changes in our aggregate indicators. In our previous paper (Kaufmann, Kraay and Mastruzzi (2005)) we showed however that this was unusual. The same is true for the latest set of changes that we calculate. We can decompose the change over time in the aggregate indicator between 1996 and 2005 for a country into two components: the change based on a common set of sources, and the remainder. We then calculate the share of the variance of changes in the aggregate indicator that is accounted for by changes in the balanced indicator. Averaging across our six indicators we find that 85 percent of the variation in changes in the aggregate indicators can be accounted for by changes in the common set of data sources, suggesting that the inclusion or exclusion of particular sources plays a relatively small role in explaining these changes.

2.4 Trends in Global Governance

We next review the available evidence on trends in global averages of governance over the expanded time period that we now cover. As we have already noted, our aggregate governance indicators are not informative about trends in global

averages because we assume that world averages of governance are zero in each period, as a choice of units. While the aggregate indicators are of course informative about the relative performance of individual (or groups of) countries, in order to assess trends in global governance we need to return to our underlying individual data sources.

In Table 5 we summarize trends in world averages in a number of our individual data sources. Most of the sources in this table are polls of experts, with data extending over the whole period 1996-2005. Other than expert polls, only one of them, GCS, is a survey with sufficiently standard format to enable comparisons over this period of time. The first column reports the number of countries covered by the source in each of the periods shown, and the next five columns present the average across all countries of each of the sources in each of the indicated 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 five columns report the standard deviation across countries for each source. The final columns reports t-statistics associated with a test of the null hypothesis that the world average score is the same in 1996 as in 2005, in 1998 as in 2005, and in 2002 as in 2005.

The picture that emerges from Table 5 is sobering. There is very little evidence of statistically significant improvements in governance worldwide. Over the longest period 1996-2005 the 18 changes reported here are divided exactly in half into 9 improvements and 9 declines in global averages, many of them quite small. There are just six cases of statistically significant changes at the 10 percent level or better (t-statistics greater than 1.64 in absolute value), and these are split between two improvements and four declines. It is not clear how much importance ought to be ascribed to these trends in world averages based on individual indicators. 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 also clear that there is substantial disagreement among sources about even the direction of changes in global averages of governance. For now we cautiously conclude that we certainly do not have any evidence of any significant improvement in governance worldwide until end-2005, and if anything the evidence is suggestive of a possible deterioration - at the very least in dimensions such as regulatory quality, rule of law, and control of corruption.

3. Are Perception Errors Correlated Among Expert Assessments of Governance?

In this section of the paper we consider two common critiques of the expert assessment individual data sources we use as inputs to our aggregate governance indicators. One concern is that the experts producing these assessments share a common set of preconceptions or prejudices about cross-country patterns of governance. If these prejudices are reflected in the assessments that they produce, this will introduce systematic errors into the data. A closely related possibility is that experts base their assessments solely on the assessments of other experts, rather than on their own view of governance in a country.¹⁵

Both of these concerns are potentially serious, because they imply that the errors made by individual sources in their estimates of governance will be correlated across sources. Intuitively this means that our observed data from these various sources will be less informative about governance than if the errors were not correlated. Moreover, these concerns have implications for how we weight different data sources when we construct our aggregate indicators. Recall that our aggregate indicators are weighted averages of the underlying individual indicators. The weights are proportional to our estimates of the precision of each indicator. We in turn infer the precision of each source from its correlation with other sources. In particular, if there are no correlations among the errors made by different sources, then sources that are more highly correlated with each other should be more precise. We thus assign greater weights to sources that tend to be highly correlated with each other.

This neat logic would however break down if we allow for the possibility that the errors made by our different sources might be correlated with each other. This is an issue that has concerned us since we began constructing governance indicators using this methodology. In Kaufmann, Kraay, and Zoido-Lobaton (1999) we showed how the standard errors of our governance estimates would increase if we simply assumed that the errors from different sources were correlated. We now examine in more detail the consequences of (a) shared prejudices, and (b) experts looking at each others' scores,

¹⁵ We do not discuss further here other potential biases in expert assessments. In Kaufmann, Kraay, and Mastruzzi (2004) we devise a test for and reject the hypothesis of ideological biases in expert assessments. In Kaufmann, Kraay, and Mastruzzi (2006) we document the relative unimportance of biases arising from "halo effects".

for our governance estimates, using two specific examples. These examples suggest to us that while these two possible sources of correlated errors are plausible *a priori*, quantitatively they do not appear to be too important in terms of their effect on our indicators.

3.1 Shared Prejudices

We first consider the case of shared prejudices. A simple way to capture this possibility is with this small variation on our basic empirical model:

$$(1) \quad y_{jk} = \alpha_k + \beta_k \cdot \left(g_j + \sigma_k \cdot \left(\sqrt{1 - \lambda_k} \cdot \varepsilon_{jk} + \sqrt{\lambda_k} \cdot p_j \right) \right)$$

where y_{jk} is the governance indicator provided by source k for country j , and g_j is the unobserved true level of governance in country j . The only difference from our basic model is that the error term now consists of a weighted average of an idiosyncratic part, ε_{jk} , and a component that is common across sources, p_j . We interpret the former as source-specific perceptions errors and assume that they are uncorrelated across sources. We interpret the latter as capturing the unobserved common "prejudices" shared by sources. For example, a common criticism is that expert assessments produced by commercial risk rating agencies focus exclusively on the business environment faced by foreign investors. We normalize p_j to have mean zero and standard deviation one. To make the problem interesting, we need to assume that p_j captures the part of respondents' prejudices that is uncorrelated with actual governance. Accordingly we assume that p_j and g_j are uncorrelated. Thus for example one can think about p_j as capturing the component of the business environment for foreign investors that is uncorrelated with true governance.¹⁶ The new parameter λ_k captures the weight placed by source k on these prejudices in coming up with its governance score for a country. Finally, we also normalize the variance of the idiosyncratic component of the error term, ε_{jk} , to have mean zero and variance one. This means that as we change the parameter λ_k , the variance of the overall error term remains constant and equal to σ_k^2 .

¹⁶ Another interpretation is the possibility of "halo" effects that we have discussed in previous work (Kaufmann, Kraay, and Mastruzzi (2005, 2006)), whereby respondents give good governance scores to rich countries simply because they are rich. Under this interpretation p_j would capture the part of halo effects that is uncorrelated with true governance.

Note also that when $\lambda_k=0$ we retrieve our original basic model in which we assume that perception errors are independent across sources and have variance equal to σ_k^2 .

In order to incorporate the effects of shared prejudices on our estimates of governance, we first need to obtain estimates of the parameter λ_k for each source, i.e. an estimate of the importance of shared prejudices in source k 's assessment of governance. Unfortunately we cannot estimate this model in general, because the statistical model in Equation (1) is not identified. Intuitively, the problem is that if we observe a high correlation between two data sources, we have no way of knowing if this is because they both place a high weight on common prejudices, i.e. they both have a high value of λ_k , or whether instead both sources have a low variance of the error term, i.e. σ_k^2 is low for the two sources.

In order to make progress we therefore need to impose some identifying assumptions. As an illustration we do this for the case of $K=3$ data sources which allows us to obtain very intuitive closed-form solutions for all of the relevant parameters as functions of the observed correlations in the data. We impose two key identifying assumptions. The first is that the variance of the overall error term is the same for all three sources, i.e. $\sigma_k^2=\sigma^2$ for $k=1,2,3$. This is mostly for convenience because it provides us with a natural benchmark: if shared prejudices play no role, then we would weight the three sources equally in our estimates of governance since the signal-to-noise ratio would be the same in all three.

The second assumption is more important. We assume that two of the three sources share some common prejudices but the third source does not. In particular we assume that $\lambda_1=\lambda_2=\lambda$ and $\lambda_3=0$. This assumption is based on the following interpretation of our two main types of data sources: expert assessments, and surveys of firms or individuals. It seems plausible that expert assessments might be prone to some form of "group-think" in which their shared preconceptions of countries affect their estimates of governance. At the same time, such "group-think" is less likely to be present in survey data where respondents are less likely to have access to the ratings provided by commercial risk rating agencies.

With these identifying assumptions in hand we can now obtain an estimate of the importance of shared perceptions based on the observed correlations in the data. Define R as the average correlation between each of the two expert assessments and the survey, i.e. $R=(R_{13}+R_{23})/2$ and define R^* as the correlation between the two expert assessments, i.e. $R^*=R_{12}$, where R_{jk} is the observed correlation between source j and source k . After some algebra (detailed in Appendix D) we can estimate the two key parameters of this extended model as:

$$(2) \quad \lambda = \frac{R^* - R}{1 - R}, \quad \sigma^2 = \frac{1 - R}{R}$$

The intuitions for these two expressions are very straightforward. The higher is R^* relative to R , i.e. the higher is the correlation among expert assessments relative to the typical correlation between an expert assessment and the survey, the higher is our estimate of λ . This is because we would attribute the relatively high correlation among expert assessments as reflecting the influence of shared prejudices on their governance ratings. The observed correlation between the expert assessments and the survey, R , provides us with information about the overall noise-to-signal ratio all three sources. The higher are these correlations, the more informative are all three data sources, i.e. the smaller is σ .

How does the presence of shared prejudices unrelated to true governance affect how we should construct our aggregate governance indicators? The simple example we have been discussing provides a useful illustration. If the expert assessments were not influenced by shared prejudices, i.e. if $\lambda=0$, then our assumption that the variance of the overall error terms is the same, i.e. $\sigma_k^2=\sigma^2$ for $k=1,2,3$ implies that we would weight all three sources equally. In the other extreme case where the error terms of the two expert assessments are perfectly correlated, i.e. if $\lambda=1$, then it is possible to show that the optimal estimate of governance would consist of a simple average of the survey and *an average of the two expert assessments*. In other words, if shared prejudices are not important, then the survey and the two expert assessments should all receive weights of one-third in the estimate of aggregate governance. If on the other hand shared prejudices dominate, then it is optimal to treat the two expert assessments as if they were just one source, and given our assumption of equal variances, to weight this one

source equally with the survey. Crucially, however, we note that the possibility of correlated errors among expert assessments does not imply that we should discard these data sources altogether. Rather, it means that we should continue to use them because even in the limiting case of perfectly correlated errors, collectively they still provide some information about unobserved governance.

This re-weighting of the survey is shown graphically in the top panel of Figure 3, which plots the weight of the survey in the aggregate indicator for alternative values of λ . When common prejudices play no role, the weight on the survey is equal to one-third. As we move to the opposite extreme where $\lambda=1$, i.e. where the error term is dominated by common prejudices, then the weight on the survey should be 0.5.¹⁷

In Table 6 we provide a sense of how important shared prejudices are quantitatively in our dataset. For our expert assessments we consider our five major data sources provided by commercial risk rating agencies (DRI, EIU, MIG, PRS, and WMO). For our survey we take GCS, which is our only very large cross-country survey of firms. In the top panel we report the two correlations of interest. We first report R^* which is the average pairwise correlation between these five expert assessments, for each of our six governance aggregates for the past four years. We also report R , which is the average pairwise correlation of each of these five expert assessments with the survey. In the bottom panel we report the estimates of λ implied by these correlations. The first thing to notice is that for four of our six indicators (PV, GE, RL and CC) R^* and R are very similar to each other, and often $R^* < R$ implying that the expert assessments are on average actually more correlated with the survey than they are with each other. This pattern of $R^* < R$ is inconsistent with shared prejudices, and so for these we have no positive estimate of λ . And in many cases R^* is only slightly larger than R , implying very small values of λ .

For two of the six aggregates, VA and RQ, we find that the expert assessments tend to be much more highly correlated with each other than with the survey, implying more substantial estimates of λ . We should however be a bit cautious in taking these

¹⁷ Clearly the importance of shared prejudices also matters for the precision of our estimates of governance. In the extreme case where $\lambda=1$ the standard error would be the same as if we had only two instead of three sources.

results at face value. Consider for example VA. The GCS survey questions we use here focus on the extent to which firms are informed by government of regulations, which is a very narrow dimension of accountability, while the expert assessments capture much broader notions of democratic accountability. This may explain the relatively low correlation of the expert assessments with the survey. To see how important this is, we took one specific question from the GCS about firms' perceptions of press freedom, and compared it with two expert assessments of press freedom (RSF, and a specific press freedom rating from FRH). As shown in the row labeled "Press Freedom", we find that the correlation between FRH and RSF is much closer to the average correlation of both of these with the survey, implying a much lower estimate of λ and suggesting a much smaller role for shared prejudices among the expert assessments.¹⁸ Overall this example suggests that the role of shared prejudices in expert assessments is at most minor.

As noted above, in this example finding positive estimates of λ as we have in some cases in Table 6 suggests that it would be appropriate to slightly decrease the weight placed on the expert assessments in our aggregate indicators. We do not make this adjustment to our benchmark governance indicators. This is because our estimate of the role of shared prejudices that we report here is very much in the spirit of an illustrative example, rather than as a definitive analysis. We have made several simplifying assumptions in order to generate these estimates as transparently as possible, and these may or may not be appropriate for our aggregate indicators that rely on our full dataset. However, we do note that a modest reduction in the weight of expert assessments in our aggregate indicators suggested by this example is unlikely to have major systematic effects on our estimates of governance. This is because, as we have mentioned earlier, the expert assessments on which we rely are in most cases quite strongly correlated with other data sources: recall that the average value of the

¹⁸ One possible objection to this exercise is that the GCS surveys firms, and it could be that such respondents share a common set of business-oriented prejudices with commercial risk rating agencies. To assess this we would need to have a large cross-country survey of individuals. Unfortunately we have only limited information to address this issue, with two of our regional surveys of individuals, Latinobarometro and Afrobarometer, covering only very small samples of countries. One possibility is to look at the Gallup survey we use that covers a somewhat larger set of 62 countries. In 2004 and 2005 we have survey questions on corruption from this source. These are typically quite highly correlated with expert assessments at about 0.7, while the expert assessments are correlated only slightly more about 0.8 in the same set of countries. This again suggests at most a modest importance of shared prejudices.

correlation of an expert assessment with the survey in Table 6 is 0.74. This means that even if we moderately shift the weights in our aggregate indicator away from the expert assessments, the resulting aggregate indicator will still be highly correlated with our benchmark assessments that do not allow for shared prejudices.

To investigate this point in more detail, we have recalculated our six governance indicators for all seven periods, weighting all of the components equally, rather than precision-weighting based on the benchmark assumption of uncorrelated errors.¹⁹ This in practice substantially reduces the weights applied to expert assessments from commercial risk rating agencies. The average (across the 42 indicators) correlation between the equally-weighted indicators and our benchmark indicators is 0.99. In only three cases is the correlation less than 0.99, and the minimum correlation is 0.97. This clearly shows that how we weight our underlying data sources does not practically affect our estimates of governance in the vast majority of cases. The main benefit however of weighting sources by their precision is that it yields somewhat smaller standard errors, allowing for more precise inference about cross-country differences and changes over time in governance. On average, the standard error of our equally-weighted indicators is about 10 percent higher than in our benchmark indicators.

3.2 Do Sources Update Based on Past Discrepancies?

We now turn to the possibility that our expert assessments at least in part base their estimates of governance on the assessments of other experts. As noted earlier, we can interpret the previous example as one case of this. Here we consider a different example in which expert assessments update their ratings based on past differences between them. In particular, suppose that we have two expert assessments that produce their estimates of governance as follows:

$$(3) \quad \begin{aligned} y_{1jt} &= g_{jt} + \lambda_1 \cdot (y_{2jt-1} - y_{1jt-1}) + \varepsilon_{1jt} \\ y_{2jt} &= g_{jt} + \lambda_2 \cdot (y_{1jt-1} - y_{2jt-1}) + \varepsilon_{2jt} \end{aligned}$$

¹⁹ In particular, we estimate the parameters of the unobserved components model imposing the restriction that the variance of the error term is the same across sources. This in turn implies an equal weighting of sources. See Equations (1)-(3) in Kaufmann, Kraay and Mastruzzi (2004).

where y_{1jt} and y_{2jt} denote the estimate of governance provided by sources 1 and 2 at time t for country j , and g_{jt} denotes the unobserved true level of governance in country j at time t . The perception errors made by the two sources consist of two parts. First, we assume that each source updates its estimates of governance in period t based on the difference between itself and the other source in the previous period. For example, if source 2 rated country j much higher than source 1 in period $t-1$, then we assume that in period t source 1 revises its estimate of governance upward, and source 2 revises its estimate of governance downward. The strength of this updating is captured by the parameters λ_1 and λ_2 . The second is an idiosyncratic component which we assume is independent across sources.²⁰

In order to assess how important this kind of updating is, we need estimates of the λ 's. One approach might be to estimate a cross-sectional regression of the assessments of each source on the lagged difference between the two sources. This however is unlikely to lead to consistent estimates of the λ 's, since the error term in such a regression, $g_{jt} + \varepsilon_{1jt}$ is likely to be correlated with the past discrepancies between sources, $y_{2jt-1} - y_{1jt-1}$, for two reasons. The first is that unobserved governance g_{jt} might be correlated with the past discrepancies between sources, $y_{2jt-1} - y_{1jt-1}$. Moreover, if the idiosyncratic component of the error terms is correlated over time, there will by construction be a correlation between ε_{1jt} and $y_{2jt-1} - y_{1jt-1}$.

A more direct approach is to look at the difference between the two sources, which is:

$$(4) \quad y_{1jt} - y_{2jt} = -(\lambda_1 + \lambda_2) \cdot (y_{1jt-1} - y_{2jt-1}) + \varepsilon_{1jt} - \varepsilon_{2jt}$$

Note that the difference between the two sources will follow an autoregressive process with a negative coefficient on the lagged difference. In particular, if source 1 rates country j higher than source 2 in the previous period, we would expect that in the current

²⁰ Note here that we have suppressed differences across sources in the α 's and β 's which capture differences in scales of measurement across sources. We do this purely for notational convenience. In the empirical application we remove the scales from our variables by expressing them as percentile ranks.

period the difference between the two sources will be smaller, if this kind of updating is important.

Equation (4) suggests two empirical tests. The first is that if this type of updating based on past discrepancies is important, we should expect to see discrepancies decline over time, and as a result, the correlations among sources should be increasing over time.²¹ This first implication can readily be checked. The top two panels of Table 7 show the pairwise correlations between our three main commercial risk rating agencies available over the full time span of our governance indicators, for 1996 and 2005. The third panel reports the change over this period in each of the pairwise correlations. These changes over time in correlations are not large, even over the decade-long period covered in the table. The median change is only 0.03, and none of the observed changes would be significant at the 90 percent confidence level. Moreover, overall the changes are quite evenly divided between seven declines and nine increases. This evidence suggests to us that there is no evidence of a systematic increase over time in the correlation among expert assessments. This in turn casts doubt on the hypothesis that these expert assessments update their estimates based on past discrepancies with each other.

Another implication of Equation (4) is that this type of updating would be a force towards creating a negative correlation over time in the pairwise differences in country assessments. We explore the empirical validity of this implication next. The simplest case to consider is when the idiosyncratic error terms in Equation (3) are independent over time. In this case, the autocorrelation over time of the differences in scores between pairs of sources is simply $-(\lambda_1 + \lambda_2)$. In this simple case we could just look at the autocorrelation of differences between sources and test whether it is negative. This would then constitute evidence of updating based on past discrepancies. The first three columns of Table 8 report the correlation over time in the pairwise difference in country rankings between 1996 and 2005, for our three main commercial risk rating agencies. The first thing to note is that these correlations are almost always positive, and average around 0.25. This does not appear to be consistent with the updating hypothesis, which

²¹ This will be true as long as the idiosyncratic error terms are not too correlated over time -- as discussed further below.

would be a force for *negative* correlations over time in pairwise differences in country rankings.

There is however a complication introduced by the possibility that the idiosyncratic error terms, ε , themselves might be correlated over time. If these error terms are very persistent, then pairwise differences in country rankings would also be very persistent, and this would obscure any negative correlation over time arising from the updating based on past discrepancies. In order to test for the importance of updating, we need to correct for this effect. After some algebra, it is possible to show that the autocorrelation of differences is given by:²²

$$(5) \quad \text{CORR}(y_{1jt} - y_{2jt}, y_{1j,t-1} - y_{2j,t-1}) = -(\lambda_1 + \lambda_2) + \rho \cdot \frac{1 - (\lambda_1 + \lambda_2)^2}{1 - \rho \cdot (\lambda_1 + \lambda_2)}$$

where ρ is the unobserved correlation over time in the idiosyncratic component of each source's error term, i.e. $\rho = \text{CORR}(\varepsilon_{1jt}, \varepsilon_{1j,t-1}) = \text{CORR}(\varepsilon_{2jt}, \varepsilon_{2j,t-1})$. This expression is quite intuitive. As already noted, if $\rho=0$, the correlation over time of the differences is just $-(\lambda_1 + \lambda_2)$. If on the other hand there is on average no updating based on past discrepancies, i.e. $\lambda_1 + \lambda_2 = 0$, then the correlation of differences over time is just the autocorrelation of the idiosyncratic errors, ρ . Finally, if the error terms are perfectly persistent, $\rho=1$, then the observed differences in scores would be perfectly correlated as well, regardless of any updating that might be present.

Given the observed correlations over time in differences between sources reported in Table 8, together with assumptions about the autocorrelation of the idiosyncratic errors, ρ , we can use Equation (5) to back out an estimate of $\lambda_1 + \lambda_2$ which summarizes the importance of updating based on lagged discrepancies between sources. We do this in the last two columns of Table 8. The fourth column reports our assumptions on ρ , the correlation over time of the idiosyncratic error terms. These are based on estimates we derived in Kaufmann, Kraay, and Mastruzzi (2005), and average

²² To derive this expression first use recursive substitution to write $y_{1jt} - y_{2jt}$ as a moving average of all past differences in the error terms, $\varepsilon_{1jt} - \varepsilon_{2jt}$. Next note that the assumptions in the text imply that $E[\varepsilon_{1jt} - \varepsilon_{2jt}, \varepsilon_{1j,t-k} - \varepsilon_{2j,t-k}] = 2\rho^k \sigma^2 / (1 - \rho^2)$. Finally use this to calculate the autocorrelations of $y_{1jt} - y_{2jt}$.

0.42 across our six indicators, indicating a substantial degree of persistence.²³ The final column reports the estimates of $\lambda_1 + \lambda_2$. Overall these estimates are quite small, averaging only 0.17. While we do not know how to attribute this total over the two sources being considered, absent better information it seems reasonable to divide it equally, implying a value of $\lambda_1 = \lambda_2 = 0.085$. This in turn implies that on average less than 10 percent of the initial difference between two expert assessments would be eliminated through this process of updating based on past discrepancies. Moreover, it is important to remember that Table 8 refers to correlations calculated over a decade between 1996 and 2005. Thus, our estimates suggest that this process of updating is so gradual that over the course of a decade it results in a reduction of only 10 percent of initial differences between sources. Considering this together with the evidence in Table 7 showing no significant trends in the pairwise correlations among expert assessments, we conclude that the evidence that expert assessments revise their country rankings based on past discrepancies with other assessments is very weak indeed.

We conclude the discussion of interdependence between expert assessments with a final example. Knack (2006) has argued that one of the expert assessments on which we rely, PRS, has in the case of corruption made systematic changes in its methodology in order to ensure that it is more correlated with the Transparency International Corruption Perceptions Index. Since the latter is an amalgam of many different assessments of corruption, including several other expert assessments, Knack (2006) argues that this is an example of one expert assessment basing its views on those of other expert assessments rather than on its own information. In particular, Knack (2006) documents that the correlation of the ICRG corruption rating with the Transparency International measure increases from 0.72 to 0.91 following what he refers to as a "massive recalibration" that occurred in October of 2001.

We do not find this critique of the PRS measure compelling because it does not appear to be systematically true. Several of the many different PRS ratings that we use have indeed had methodological breaks in the past, like the one for corruption that Knack (2006) focuses on. If the objective of such breaks is to generate new ratings that

²³ These estimates are only suggestive, as they are not fully consistent with the statistical model considered here. In particular, they are based on a model which does not allow for error terms to be correlated across sources. It is not clear whether this restriction would bias our estimates of ρ up or down.

are more correlated those of with other experts, as suggested by Knack (2006), then we should systematically expect increases in correlations with other experts when comparing the period before and after the methodological break. In contrast, we should see no change in the correlation of PRS with other expert assessments for series that did not have methodological breaks.

In Table 9 we examine the 10 governance indicators produced by PRS that we use as an input to our aggregate governance indicators. For each of these series we look at the number of countries whose scores change from month to month on the monthly PRS data. We identify series with methodological breaks by looking for sharp increases in the number of countries whose scores change in a particular month. This method clearly turns up cases of methodological breaks. In the case of corruption, for example, in a typical month only 1.4 percent of countries in the PRS sample change scores relative to the previous month, while in October of 2001 41 percent of the countries in the sample change scores. As shown in Table 9, the main methodological breaks occur in 1997 and in 2001. Importantly for us, not all series have breaks in both periods. In 1997 for example we detect methodological breaks in the PRS series "Democratic Accountability" which we map to VA, and "Bureaucratic Quality" which we map to GE, while the remaining 8 PRS series do not have breaks in 1997. Similarly in 2001 there are five series with breaks and five series without breaks.

In the columns of Table 9 we report the correlations of the PRS series with our two other major expert assessments, DRI and EIU, in the years before and after the break. For corruption, for example, we report the correlation of PRS rating with the corruption ratings provided by DRI and EIU, in 2000 and in 2002. We do this for all 10 PRS indicators, separating those with and without methodological breaks. We then ask whether there is any evidence that the PRS indicators with methodological breaks are more likely to see their correlations with other expert assessments increase than those series without breaks. The answer appears to be no. Consider for example the comparison of correlations with PRS in 2000 and 2002. For the five PRS series with methodological breaks, the median change in the correlation of each of these with DRI and EIU is just 0.01, and in only about half of the cases is the change in correlation positive. The median change in the correlation with DRI and EIU for series without methodological breaks is almost exactly the same, at -0.01. Similarly in 1997, we do not

see much in the way of systematic evidence that PRS series with methodological breaks are more likely to become more correlated with other expert assessments. Based on this systematic look at the behavior of the PRS indicators around the time of methodological breaks, we do not find it plausible to argue as does Knack (2006) that PRS has exploited methodological breaks in the past to increase their correlation with the ratings of other expert assessments.

4. Conclusions

In this paper we have reported on the latest update of our aggregate governance indicators. With this update we have (a) moved to an annual frequency for reporting, (b) for the first time made available virtually all of the individual indicators underlying the aggregate indicators, and (c) provided new evidence on the reliability of expert assessments of governance that form part of our aggregate indicators. It is our hope that more timely annual reporting as well as access to individual indicators will make the aggregate indicators more useful to users in academic and policymaking circles.

We nevertheless emphasize to all users the limitations of these measures of governance, which are shared by virtually all efforts to measure governance across countries and over time. The aggregate indicators we construct are useful for broad cross-country and over time comparisons of governance, but all such comparisons should take appropriate account of the margins of error associated with the governance estimates. These margins of error are not unique to our perceptions-based measures but are present -- if not explicitly acknowledged -- in any effort to measure governance. They naturally reflect the inherent difficulty in measuring something as complicated and multifaceted as governance. However, we have shown the feasibility of using the aggregate indicators to make comparisons of governance across countries and over time, subject to appropriate consideration of margins of error. In fact, for 2005 we have seen that fully 60% of all cross-country comparisons result in highly-significant differences, and that nearly one-third of countries have experienced substantial changes in at least one dimension of governance between 1996 and 2005.

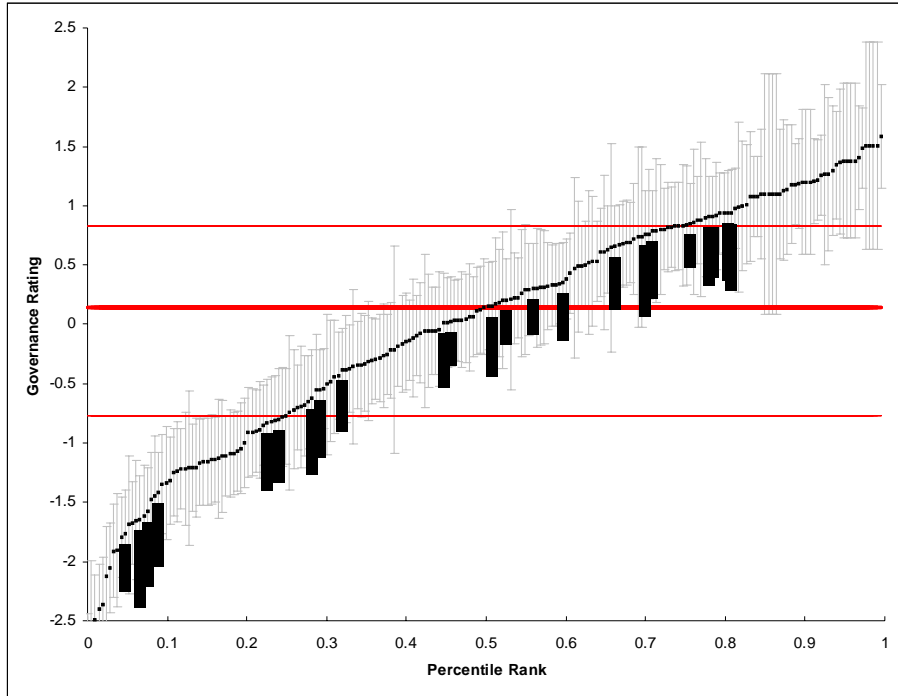
We also caution users that the aggregate indicators can in some circumstances be a rather blunt tool for policy advice at the country level. We expect that the provision of the underlying data will help users in identifying -- and acting upon -- more specific aspects of governance that may be problematic in a given country. And we also encourage using these aggregate and individual indicators in conjunction with a wealth of possible more detailed and nuanced sources of country-level data on governance in formulating policy advice.

References

- Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton (1999a). "Aggregating Governance Indicators." World Bank Policy Research Working Paper No. 2195, Washington, D.C.
- (1999b). "Governance Matters." World Bank Policy Research Working Paper No. 2196, Washington, D.C.
- (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 (2006). "Measuring Governance Using Perceptions Data", in Susan Rose-Ackerman, ed. Handbook of Economic Corruption. Edward Elgar.
- Knack, Steven (2006). "Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators". World Bank Policy Research Working Paper 3968.

Figure 1: Margins of Error for Governance Indicators, 2005

Political Stability and Absence of Violence



Control of Corruption

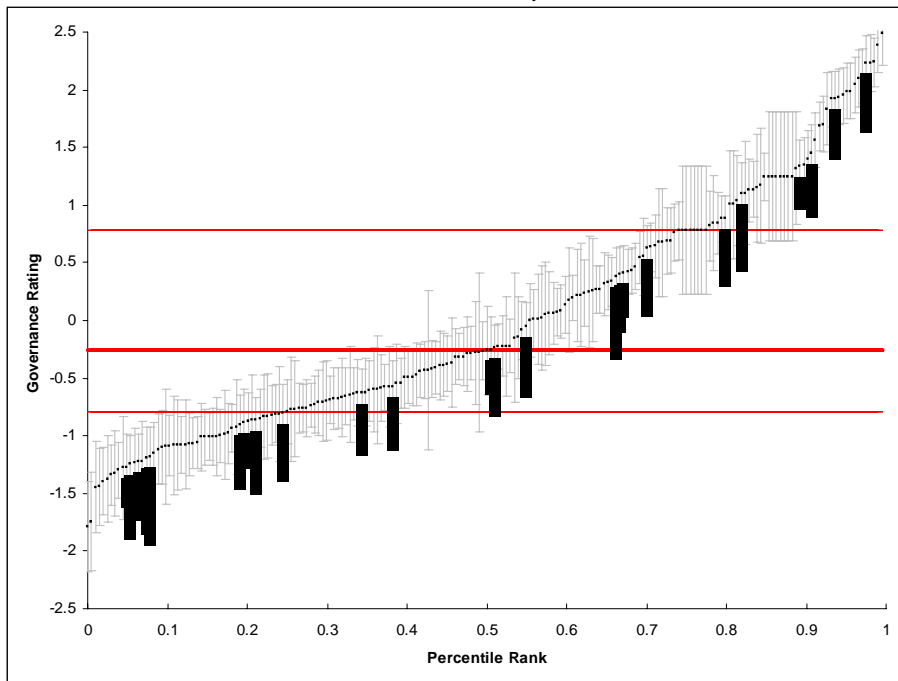
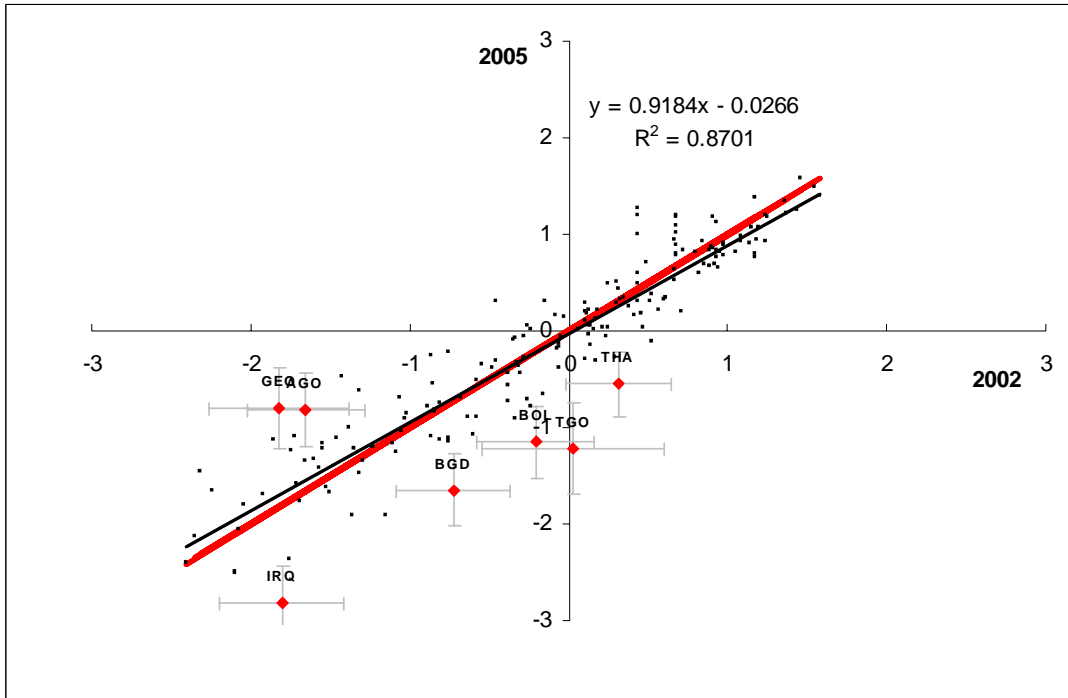


Figure 2: Changes Over Time in Governance Indicators 2002-2005

Political Stability and Absence of Violence



Control of Corruption

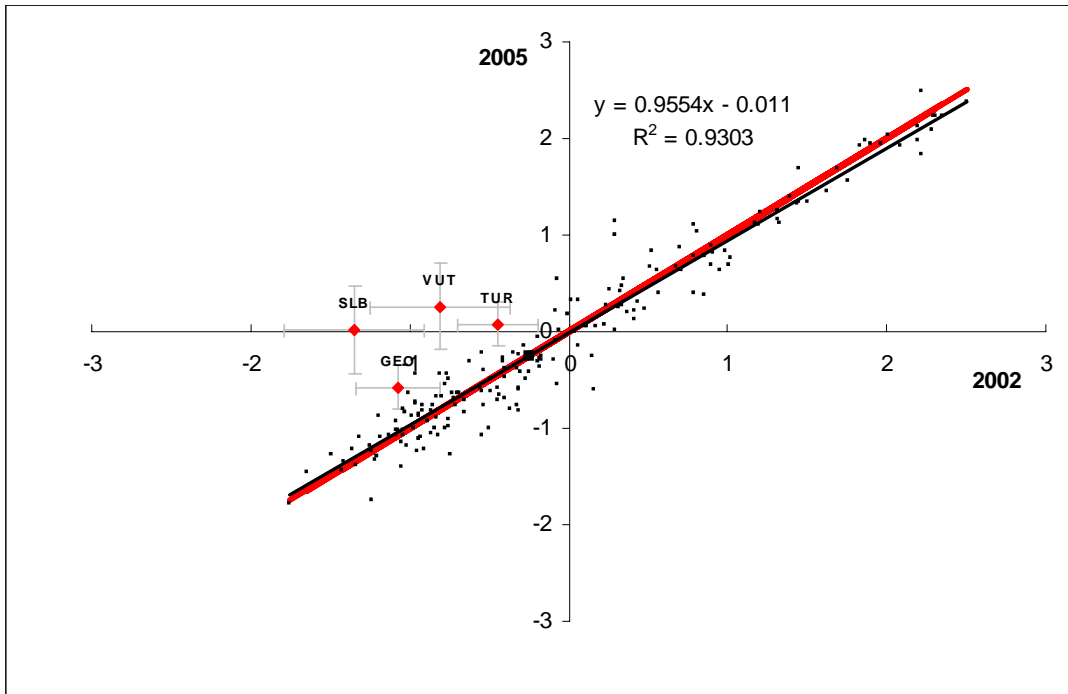


Figure 3: Reweighting Survey Due to Shared Prejudices

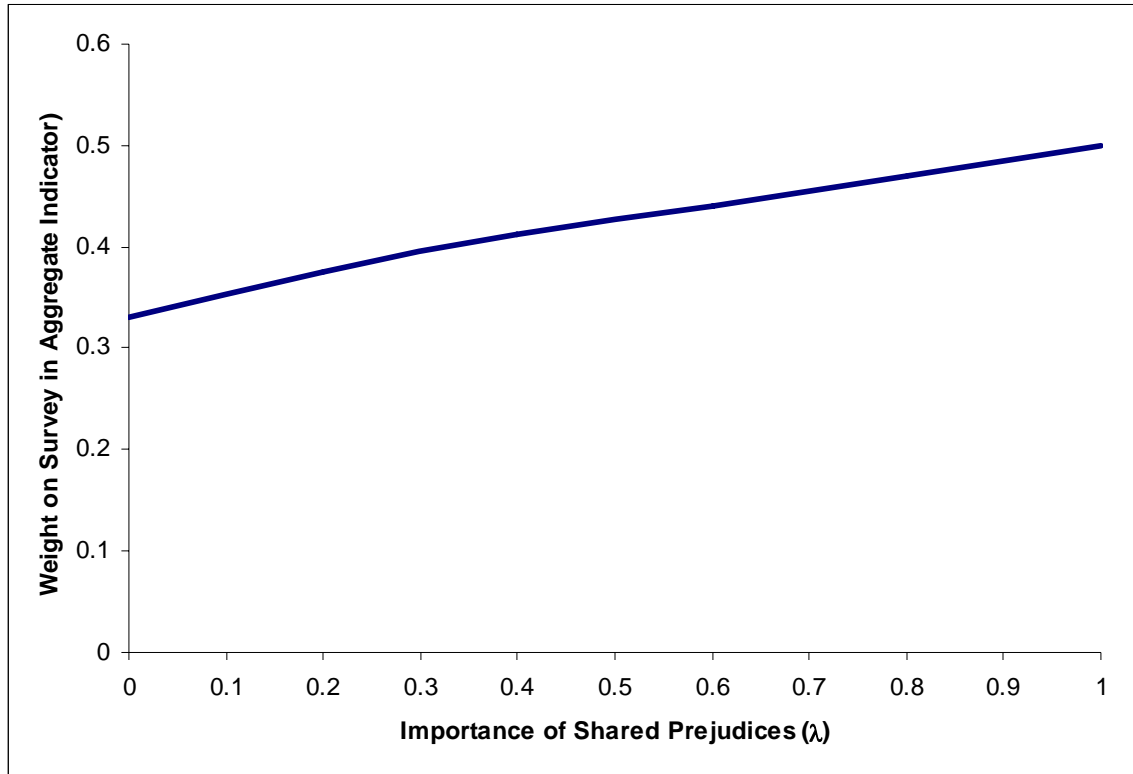


Table 1: Sources of Governance Data

Source	Publication	Code	Type	Public	Country Coverage	Representative	1996	1998	2000	2002	2003	2004	2005
African Development Bank	Country Policy & Institutional Assessments	ADB	Poll	No	50				x	x	x	x	x
Afrobarometer	Afrobarometer Survey	AFR	Survey	Yes	18				x	x	x	x	x
Asian Development Bank	Country Policy & Institutional Assessments	ASD	Poll	Partial	26				x	x	x	x	x
Bertelsmann Foundation	Bertelsmann Transformation Index	BTI	Poll	Yes	119							x	x
Brown University's Center for Public Policy	Global E-Governance	EGV	Poll	Yes	192	x				x	x	x	x
Business Environment Risk Intelligence	Business Risk Service	BRI	Poll	Yes	50		x	x	x	x	x	x	x
Business Environment Risk Intelligence	Qualitative Risk Measure	QLM	Poll	Yes	115	x	x	x	x	x	x	x	x
Columbia University	State Capacity Project	CDU	Poll	Yes	108	x			x	x	x	x	
Economist Intelligence Unit	Country Risk Service	EIU	Poll	Yes	120	x	x	x	x	x	x	x	x
European Bank for Reconstruction & Development	Transition Report	EBR	Poll	Yes	27		x	x	x	x	x	x	x
Freedom House	Countries at the Crossroads	CCR	Poll	Yes	30							x	x
Freedom House	Nations in Transition	FHT	Poll	Yes	27		x	x	x	x	x	x	x
Freedom House	Freedom in the World	FRH	Poll	Yes	192	x	x	x	x	x	x	x	x
Gallup International	Voice of the People Survey	GAL	Survey	Yes	69					x	x	x	x
Global Insight	Global Risk Service	DRI	Poll	Yes	111	x	x	x	x	x	x	x	x
Global Insight	Business Conditions and Risk	WMO	Poll	Yes	202	x				x	x	x	x
Heritage Foundation	Economic Freedom Index	HER	Poll	Yes	161	x	x	x	x	x	x	x	x
IJET Travel Intelligence	Country Security Risk Assessment	IJT	Poll	Yes	167	x						x	x
Institute for Management and Development	World Competitiveness Yearbook	WCY	Survey	Yes	49		x	x	x	x	x	x	x
International Research & Exchanges Board	Media Sustainability Index	MSI	Poll	Yes	19					x	x	x	x
Latinobarometro	Latinobarometro Surveys	LBO	Survey	Yes	18		x	x	x	x	x	x	x
Merchant International Group	Grey Area Dynamics	MIG	Poll	Yes	159	x				x	x	x	x
Political & Economic Risk Consultancy	Corruption Survey	PRC	Survey	Yes	10				x	x	x	x	x
Political Risk Services	International Country Risk Guide	PRS	Poll	Yes	140	x	x	x	x	x	x	x	x
Reporters Without Borders	Reporters Without Borders	RSF	Poll	Yes	165	x				x	x	x	x
State Department	Trafficking in People Report	TPR	Poll	Yes	149	x			x	x	x	x	x
State Department / Amnesty International	Human Rights Dataset	HUM	Poll	Yes	192	x	x	x	x	x	x	x	x
World Bank	Business Enterprise Environment Survey	BPS	Survey	Yes	27				x	x	x	x	x
World Bank	World Business Environment Survey	WBS	Survey	Yes	80	x			x	x			
World Bank	Country Policy & Institutional Assessments	PIA	Poll	Partial	136		x	x	x	x	x	x	x
World Economic Forum	Global Competitiveness Report	GCS	Survey	Yes	117	x	x	x	x	x	x	x	x

Table 2: Summary Statistics on Governance Indicators

	<u>Voice and Accountability</u>	<u>Political Stability</u>	<u>Government Effectiveness</u>	<u>Regulatory Quality</u>	<u>Rule of Law</u>	<u>Control of Corruption</u>	<u>Overall</u>
Number of Countries							
1996	193	179	181	183	168	152	176
1998	192	166	184	185	186	184	183
2000	192	166	187	188	188	187	185
2002	199	186	202	197	197	197	196
2003	201	186	202	197	197	197	197
2004	207	207	208	204	208	204	206
2005	208	213	210	203	208	204	208
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	6	5	9	7	6
2002	7	7	8	7	11	8	8
2003	7	7	8	7	11	8	8
2004	8	8	9	8	12	8	9
2005	8	7	9	8	11	8	9
Proportion of Countries with Only One Data Source							
1996	15	18	22	15	7	18	16
1998	14	10	19	13	11	18	14
2000	14	6	7	7	5	7	8
2002	10	10	5	7	7	8	8
2003	10	9	5	7	7	7	7
2004	6	6	8	8	8	8	7
2005	6	7	9	8	8	7	7
Average Standard Error							
1996	0.28	0.40	0.30	0.37	0.28	0.34	0.33
1998	0.26	0.32	0.36	0.41	0.27	0.29	0.32
2000	0.26	0.34	0.27	0.40	0.23	0.26	0.29
2002	0.22	0.27	0.23	0.23	0.19	0.21	0.23
2003	0.19	0.28	0.23	0.21	0.18	0.20	0.22
2004	0.18	0.26	0.23	0.22	0.19	0.21	0.21
2005	0.17	0.28	0.20	0.20	0.19	0.19	0.21

Table 3: Large Changes in Governance, 2002-2005

	Governance Score			Agree	No change	Dis-agree	Agree/ (agree+ Disagree)	Sources Added	Sources Subtracted
	2005	2002	Change						
Voice & Accountability									
UNITED ARAB EMIRATES	-1.08	-0.49	-0.59	4	1	1	0.80	2	0
RUSSIA	-0.85	-0.45	-0.40	8	0	3	0.73	2	1
SINGAPORE	-0.29	0.50	-0.79	4	1	2	0.67	3	0
IRAQ	-1.47	-2.15	0.69	6	0	0	1.00	1	0
KENYA	-0.12	-0.69	0.57	5	1	1	0.83	3	1
LIBERIA	-0.92	-1.57	0.65	5	0	0	1.00	1	1
UKRAINE	-0.26	-0.68	0.43	6	2	1	0.86	2	1
SERBIA AND MONTENEGRO	0.12	-0.30	0.42	7	1	0	1.00	3	0
Political Stability									
BANGLADESH	-1.65	-0.73	-0.92	4	0	3	0.57	1	1
BOLIVIA	-1.15	-0.21	-0.95	5	0	2	0.71	0	0
IRAQ	-2.82	-1.80	-1.01	4	1	1	0.80	1	0
TOGO	-1.22	0.03	-1.24	3	0	0	1.00	2	0
THAILAND	-0.55	0.31	-0.86	7	0	2	0.78	1	1
ANGOLA	-0.82	-1.66	0.84	6	0	0	1.00	1	1
GEORGIA	-0.80	-1.83	1.03	2	1	1	0.67	2	1
Government Effectiveness									
ZIMBABWE	-1.42	-0.81	-0.61	6	0	3	0.67	2	0
* ST. KITTS AND NEVIS	1.00	-0.27	1.28	0	0	2	0.00	1	0
* ST. VINCENT AND THE GRENADINES	1.07	-0.20	1.27	1	0	1	0.50	1	0
SLOVAK REPUBLIC	0.95	0.46	0.49	9	1	0	1.00	2	0
Regulatory Quality									
VENEZUELA	-1.15	-0.57	-0.58	6	1	2	0.75	1	0
* VANUATU	0.05	-1.23	1.29	2	0	0	1.00	1	0
IRAQ	-1.61	-2.30	0.69	4	0	1	0.80	0	0
Control of Corruption									
GEORGIA	-0.57	-1.07	0.50	5	0	2	0.71	2	1
TURKEY	0.08	-0.45	0.52	8	2	1	0.89	2	1
* SOLOMON ISLANDS	0.02	-1.35	1.37	2	0	0	1.00	1	0
* VANUATU	0.26	-0.81	1.07	1	1	0	1.00	1	0

* 2005 estimates based on fewer than three data sources.

Table 4: Agreement Ratio for Changes in Governance, 2002-2005

ALL CHANGES					
	Sample	<u>Agree</u>	<u>No Change</u>	<u>Disagree</u>	<u>Agree / (Agree + Disagree)</u>
Voice and Accountability	199	2.28	1.32	1.35	0.63
Political Stability	186	2.17	0.73	1.25	0.63
Government Effectiveness	196	2.56	1.30	1.74	0.60
Regulatory Quality	196	2.54	0.97	1.54	0.62
Rule of Law	196	2.62	3.22	1.53	0.63
Control of Corruption	196	2.18	2.17	1.16	0.65
Average	195	2.4	1.6	1.4	0.63
SIGNIFICANT CHANGES (90%)					
	Sample	<u>Agree</u>	<u>No Change</u>	<u>Disagree</u>	<u>Agree / (Agree + Disagree)</u>
Voice and Accountability	8	5.6	0.8	1.0	0.88
Political Stability	7	4.4	0.3	1.3	0.94
Government Effectiveness	4	4.0	0.3	1.5	0.94
Regulatory Quality	3	4.0	0.3	1.0	0.92
Rule of Law	0
Control of Corruption	4	4.0	0.8	0.8	0.84
Average	4	4.4	0.5	1.1	0.91

Table 5: Global Trends in Governance 1996-2005 for Selected Sources

	[Quasi Balanced Sample]* ** ***											t-statistic 1996-2005	t-statistic 1998-2005	t-statistic 2002-2005
	World Average					Std. Dev. Across Countries								
	Sample	1996	1998	2000	2002	2005	1996	1998	2000	2002	2005			
Voice and Accountability														
EIU	115	0.41	0.42	0.42	0.46	0.45	0.30	0.32	0.31	0.28	0.28	1.15	0.89	-0.19
FRH	190	0.562	0.565	0.569	0.582	0.59	0.31	0.31	0.30	0.30	0.30	1.33	0.90	0.38
GCS (Press Freedom / Parliament) **	97	0.57	0.55	0.15	0.15	-0.81
PRS *	140	0.63	0.63	0.63	0.63	0.67	0.25	0.27	0.27	0.26	0.26	1.26	1.34	1.18
WMO	186	0.55	0.57	0.26	0.25	0.53
Political Stability														
EIU	115	0.54	0.51	0.56	0.54	0.56	0.29	0.30	0.30	0.28	0.27	0.53	1.33	0.43
GCS (cost of terrorism) **	97	0.66	0.67	0.17	0.14	0.44
PRS *	140	0.70	0.76	0.78	0.75	0.74	0.13	0.14	0.11	0.12	0.11	2.54	-1.79	-0.94
WMO	186	0.67	0.67	0.24	0.22	0.10
Government Effectiveness														
EIU	115	0.39	0.45	0.44	0.38	0.37	0.30	0.24	0.24	0.29	0.31	-0.46	-2.33	-0.40
GCS (infrastructure quality) **	97	0.49	0.52	0.54	0.53	0.50	0.22	0.21	0.25	0.24	0.23	0.55	-0.53	-0.80
PRS *	140	0.58	0.51	0.54	0.54	0.53	0.24	0.21	0.29	0.29	0.28	-1.47	0.80	-0.24
WMO	186	0.56	0.57	0.23	0.23	0.36
Regulatory Quality														
EIU	115	0.42	0.51	0.54	0.25	0.25	0.24	3.77	..	1.08
GCS (burden of regulations) **	97	0.44	0.48	0.42	0.30	0.34	0.15	0.15	0.12	0.13	0.11	-5.17	-7.59	2.60
HERITAGE ***	155	0.50	0.48	0.49	0.50	0.51	0.30	0.31	0.31	0.29	0.18	0.39	0.94	0.23
WMO	186	0.58	0.59	0.25	0.25	0.35
Rule of Law														
EIU	115	0.47	0.50	0.48	0.52	0.51	0.27	0.30	0.30	0.26	0.26	0.96	0.36	-0.38
GCS (cost of organized crime / quality of police / independent judiciary) **	97	..	0.65	0.62	0.56	0.55	..	0.21	0.23	0.21	0.19	..	-3.45	-0.43
HERITAGE ***	155	0.50	0.48	0.46	0.44	0.44	0.30	0.31	0.31	0.29	0.28	-1.87	-1.12	-0.18
PRS *	140	0.72	0.62	0.65	0.62	0.64	0.23	0.26	0.23	0.24	0.21	-3.31	0.50	0.75
QLM	115	0.45	0.45	0.46	0.46	0.44	0.29	0.29	0.30	0.30	0.30	-0.07	-0.23	-0.34
WMO	186	0.58	0.59	0.23	0.23	0.38
Control of Corruption														
EIU	115	0.35	0.34	0.33	0.35	0.33	0.31	0.32	0.31	0.32	0.35	-0.29	-0.23	-0.33
GCS (bribe frequency)**	97	..	0.66	0.69	0.64	0.63	..	0.29	0.25	0.22	0.18	..	-0.86	-0.35
PRS *	140	0.59	0.51	0.47	0.41	0.41	0.21	0.21	0.21	0.19	0.20	-7.08	-3.85	0.33
QLM	115	0.39	0.40	0.40	0.40	0.38	0.29	0.29	0.29	0.29	0.29	-0.35	-0.54	-0.51
WMO	186	0.52	0.53	0.27	0.26	0.32

* PRS Country coverage in 1996: 130, all other periods 140.

** GCS Country coverage in 1996: 58; in 1998: 59; in 2000: 75; 2002-80; 2003-2005: 97.

*** Heritage Country coverage in 1996: 137; all other periods 155.

Table 6: Shared Prejudices in Expert Assessments

<i>Correlations Among Experts (R*) and Correlations with Survey (R)</i>					
		<u>2002</u>	<u>2003</u>	<u>2004</u>	<u>2005</u>
Voice and Accountability	R*	0.85	0.85	0.86	0.84
	R	0.66	0.65	0.61	0.74
Political Stability	R*	0.74	0.78	0.79	0.76
	R	0.73	0.76	0.78	0.80
Government Effectiveness	R*	0.74	0.78	0.79	0.76
	R	0.73	0.76	0.78	0.80
Regulatory Quality	R*	0.68	0.72	0.73	0.72
	R	0.52	0.61	0.64	0.60
Rule of Law	R*	0.75	0.78	0.80	0.78
	R	0.79	0.78	0.80	0.81
Control of Corruption	R*	0.77	0.80	0.82	0.80
	R	0.79	0.83	0.83	0.82
Press Freedom	R*	0.79	0.83	0.82	0.83
	R	0.80	0.79	0.79	0.77
<i>Implied Estimates of λ</i>					
		<u>2002</u>	<u>2003</u>	<u>2004</u>	<u>2005</u>
Voice and Accountability		0.57	0.56	0.64	0.40
Political Stability		0.03	0.09	0.05	..
Government Effectiveness		0.03	0.09	0.05	..
Regulatory Quality		0.32	0.28	0.26	0.30
Rule of Law	
Control of Corruption	
Press Freedom		..	0.17	0.16	0.28

Note: ".." indicates that estimates of R* and R are not consistent with a positive value of λ

Table 7: Trends in Correlations Among Expert Assessments

	1996			
	dri-eiu	dri-prs	eiu-prs	
Voice and Accountability	0.82	
Political Stability	0.71	0.59	0.66	
Government Effectiveness	0.81	0.67	0.79	
Regulatory Quality	0.70	0.56	0.55	
Rule of Law	0.82	0.72	0.72	
Control of Corruption	0.83	0.65	0.75	
	2005			
	dri-eiu	dri-prs	eiu-prs	
Voice and Accountability	0.80	
Political Stability	0.81	0.62	0.54	
Government Effectiveness	0.74	0.62	0.84	
Regulatory Quality	0.64	0.79	0.78	
Rule of Law	0.84	0.70	0.77	
Control of Corruption	0.81	0.70	0.84	
	Change			
	dri-eiu	dri-prs	eiu-prs	
Voice and Accountability	-0.02	
Political Stability	0.10	0.03	-0.12	
Government Effectiveness	-0.07	-0.06	0.05	
Regulatory Quality	-0.06	0.23	0.23	
Rule of Law	0.02	-0.02	0.05	
Control of Corruption	-0.03	0.06	0.09	
	Summary of Changes			
	dri-eiu	dri-prs	eiu-prs	overall
Median Change	-0.03	0.03	0.05	0.03
Changes<0	3	2	2	7
Changes>0	2	3	4	9

Note: Sample size = 103 for VA and 87 for remaining indicators.
This implies standard errors for changes of 0.14 and 0.15.

Table 8: Serial Correlation of Pair-wise Differences in Country Rankings

	Difference in Ranks (Correlation of 1996 with 2005)			average R	Average ρ	Implied $(\lambda_1+\lambda_2)$
	dri-eiu	dri-prs	eiu-prs			
Voice and Accountability			0.32	0.32	0.51	0.21
Political Stability	0.15	0.42	0.52	0.36	0.40	0.04
Government Effectiveness	0.21	0.23	0.09	0.18	0.35	0.18
Regulatory Quality	0.28	0.10	-0.06	0.11	0.33	0.23
Rule of Law	0.35	0.58	0.31	0.41	0.54	0.16
Control of Corruption	0.19	0.34	0.10	0.21	0.38	0.18
Average	0.24	0.33	0.21	0.27	0.42	0.17

Table 9: Methodological Breaks in PRS Indicators

	1996	1996	1998	1998	Change	Change	
	DRI	EIU	DRI	EIU	DRI	EIU	
Series With Breaks in 1997							
Democratic Accountability (VA)	..	0.83	..	0.75	..	-0.08	
Bureaucratic Quality (GE)	0.67	0.79	0.68	0.62	0.01	-0.18	
<i>Median Change</i>							-0.08
<i>Fraction Increase</i>							0.33
Series Without Breaks in 1997							
Military in Politics (VA)		0.72		0.73	..	0.02	
Internal Conflict (PV)	0.62	0.60	0.74	0.77	0.12	0.17	
External Conflict (PV)	0.33	0.43	0.51	0.51	0.19	0.08	
Ethnic Tensions (PV)	0.50	0.56	0.61	0.66	0.11	0.10	
Government Stability (PV)	0.04	0.14	0.02	0.10	-0.02	-0.04	
Investment Profile (RQ)	0.56	..	0.54	..	-0.02	..	
Law and Order (RL)	0.72	0.71	0.74	0.77	0.03	0.05	
Corruption (CC)	0.64	0.75	0.69	0.68	0.04	-0.07	
<i>Median Change</i>							0.05
<i>Fraction Increase</i>							0.71
	2000	2000	2002	2002	Change	Change	
	DRI	EIU	DRI	EIU	DRI	EIU	
Series With Breaks in 2001							
Internal Conflict (PV)	0.68	0.78	0.76	0.75	0.08	-0.03	
External Conflict (PV)	0.50	0.51	0.44	0.48	-0.06	-0.03	
Government Stability (PV)	0.09	0.19	0.19	0.20	0.10	0.01	
Investment Profile (RQ)	0.55		0.68		0.13		
Corruption (CC)	0.75	0.75	0.74	0.84	-0.02	0.09	
<i>Median Change</i>							0.01
<i>Fraction Increase</i>							0.56
Series Without Breaks in 2001							
Military in Politics (VA)		0.71		0.70		-0.01	
Democratic Accountability (VA)		0.75		0.79		0.05	
Ethnic Tensions (PV)	0.46	0.52	0.37	0.44	-0.09	-0.08	
Bureaucratic Quality (GE)	0.75	0.82	0.73	0.86	-0.02	0.04	
Law and Order (RL)	0.74	0.70	0.71	0.70	-0.03	0.00	
<i>Median Change</i>							-0.01
<i>Fraction Increase</i>							0.38

Note: Break in series identified by fraction of countries that change from month to month in PRS

Series identified as having breaks have 40 percent or more of countries change in a single month

Series identified as having no breaks have less than 18 percent of countries change in a single month.

Correlations based on common sample of 97 countries