Chapter 2

Measuring governance using cross-country perceptions data

Daniel Kaufmann, Aart Kraay, and Massimo Mastruzzi*

'I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.'
'If you cannot measure it, you cannot improve it.'

-- Sir William Thomas Kelvin

Today there is widespread consensus among policymakers and academics that good governance and strong institutions lie at the core of economic development. The intellectual foundations for this view are not new, and go back at least to the seminal work of Douglass North and earlier. What is new is that over the past 10 years there has been an explosion of careful empirical work that has documented a strong causal link running from better institutions to better development outcomes. Figure 2.1 summarizes the main results from several recent cross-country empirical studies. On the horizontal axis we graph a measure of institutional quality capturing the protection of property rights (the Rule of Law indicator is described in more detail below). On the vertical axis we plot real GDP per capita, and we have normalized both variables to have a mean of zero and a standard deviation of one. The country-level data in the graph illustrates the strong correlation between governance and per capita incomes. This recent research has gone beyond the simple correlation shown in the graph to identify a strong causal impact

of governance on development. The upward-sloping lines capture several estimates of the causal impact of governance on per capita incomes that have been isolated in recent studies[1] using various techniques. The striking observation that emerges from this graph is that the estimated causal impact of institutions on economic development is large: a realistic one-standard deviation improvement in governance would raise per capita incomes in the long run by a factor of two to three. Such improvement in governance corresponds, for instance, to the improvement from the levels of Somalia to those of Laos, or from Laos to Lebanon, or from that of Lebanon to Italy, or from Italy to Canada. [figure 2.1]

A key factor enabling this line of recent research and informing policy discussions related to governance has been the availability of more and better cross-country and within-country data on governance and institutional quality. One such measurement effort has been our work since the late 1990s to construct a dataset of aggregate cross-country governance indicators using subjective data on perceptions of governance from a large number of data sources. In Section 2 of this paper we report on the latest update of our governance indicators, which measure six dimensions of governance over the period 1996-2004 and spanning 209 countries and territories. The indicators are based on several hundred individual variables measuring perceptions of governance, drawn from 37 separate data sources constructed by 31 different organizations.

Reformers in many governments, aid donors, members of civil society, and investors increasingly recognize governance as key for development. This in turn has increased the demand for monitoring the quality of governance both across countries and within countries over time. For example, one of the eligibility criteria for the United

States governments new aid program, the Millennium Challenge Account (MCA), is that a country must score above the median of all potentially-eligible countries on the Control of Corruption indicator described in this paper.[2] One of the messages from our work is that it is important when employing such measures to take into account the inevitable uncertainty associated with estimates of governance. An attractive feature of our approach to measuring governance is that it allows us to quantify the precision and reliability of our estimates of governance. Over time the addition of data has improved the precision of our governance indicators. However, the margins of error associated with estimates of governance are not trivial, and need to be taken into account when comparing governance across countries.

The same margins of error also complicate the measurement of changes over time in governance, an issue of obvious concern to many policymakers. In Section 3 we present new results on how to assess the statistical significance of changes over time in our measures of governance. We find that although many of the observed changes over time in our governance indicators are too small to signal statistically or economically meaningful changes in governance, there are countries where there have been substantial changes in governance, both improvements and declines. We also find that the likelihood of observing significant changes increases substantially with the length of the time period under consideration. Importantly, in examining some of our underlying data sources we also find that there is no evidence of changes in global averages of governance worldwide. Although our aggregate indicators are scaled to have the same mean and standard deviation in each period and thus only track relative changes in governance over

time, the absence of trends in global averages suggests that there is little difference between these relative and absolute changes in governance.

In Section 4 we discuss several issues that arise when using perceptions-based data to measure governance across countries. We first note that often subjective data is the only type of information available for various dimensions of governance, and that the quality of subjective data on governance has improved over time. We also note that the margins of error we emphasize in our work are not unique to the perceptions data we use to construct our aggregate governance indicators: measurement error is pervasive among all measures of governance and institutional quality. An advantage of our measures of governance is that we are able to be explicit about the accompanying margins of error, whereas these are most often left implicit with objective measures of governance. To remedy this we provide a simple calculation which suggests that margins of error in objective indicators of governance are at least as large as those we report for our subjective indicators. We also investigate in more detail discrepancies between subjective and objective measures of very specific dimensions of the regulatory environment. We show that firms' survey responses about their tax burden and the ease of starting a new business reflect not only the *de jure* regulations governing these issues, but also the overall institutional and governance environment in which these regulations are applied. Finally, we show that concerns about the importance of ideological biases in subjective governance assessments are relatively unimportant. These findings emphasize the importance of relying on a full range of measures of governance, and not exclusively on either subjective or objective measures, when assessing the quality of governance across countries.

We began by noting that there is widespread consensus among academics and policymakers that governance is important for economic development. But this view is not without its critics. In Section 5 we address two prominent lines of such criticism. The first argues that the strong positive correlation observed between subjective measures of governance and per capita incomes does not reflect a causal impact of governance on development, but rather is mostly due to "halo effects" – respondents rating countries might provide good governance scores to richer countries simply because they are richer. While this is certainly a possible source of bias, we show that it is unlikely to lead to a significant upward bias in the correlation between income and governance. The second line of criticism is implicitly based on the view that the observed correlation between governance and per capita income largely reflects and important causal effect running from incomes to governance: as countries get richer, institutional quality will improve. This view has led some observers of the poor development performance of countries in sub-Saharan Africa to argue that the on average poor governance of countries in the region should be "discounted" because per capita incomes in the region are also low. However, we argue that existing evidence does not support a strong causal channel operating in this direction – most of the correlation between governance and per capita incomes reflects causation from governance to per capita incomes. In light of this we suggest that it would be inappropriate to divert attention from the weak average governance performance of the region simply because the region is poor. While we focus on Africa because of the recent emphasis in the aid community on the region, the fallacy of discounting the extent of misgovernance in a country or region due to low incomes applies more generally to any setting with poor governance and low incomes.

1. Updated governance indicators for 1996-2004

In this section we briefly describe the update of our governance indicators for 2004, as well as some minor backwards revisions to the indicators for 1996-2002. Our basic methodology has not changed from past years, and a detailed discussion can be found in Kaufmann, Kraay, and Mastruzzi (2004), and in the working paper version of this chapter (Kaufmann, Kraay, and Mastruzzi 2005). We construct measures of six dimensions of governance:

1. *Voice and Accountability* – measuring political, civil, and human rights
2. *Political Instability and Violence* – measuring the likelihood of violent threats to—or changes in—government, including terrorism.
3. *Government Effectiveness* – measuring the competence of the bureaucracy and the quality of public service delivery.
4. *Regulatory Burden* – measuring the incidence of market-unfriendly policies.
5. *Rule of Law* – measuring the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence.
6. *Control of Corruption* – measuring the exercise of public power for private gain, including both petty and grand corruption, or "state capture".

Our estimates of governance are based on a large number of individual data sources 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. For the 2004 round of the governance indicators, we rely on a total of 352 individual variables measuring different dimensions of governance. These are taken from 37 different sources produced by 31 different organizations. A full list of the data sources, as well as a detailed description of how individual perceptions measures are assigned to our six dimensions of governance, can be found in Kaufmann, Kraay, and Mastruzzi (2005).

Our data sources reflect the views of a very diverse group of respondents. Several of our data sources 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 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, the UN Economic Commission for Africa, and the World Bank), reflecting these individuals' in-depth knowledge of 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 EIU and DRI) base their assessments on a global network of correspondents typically living in the country they are rating.

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? We approach this question using a statistical method known as an unobserved components model, which allows us to extract the common dimension of unobserved governance from the many individual data sources at our disposal. Details on this statistical approach can be found in Kaufmann, Kraay, and Mastruzzi (2005). 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 this chapter, this imprecision is not a consequence of our reliance on subjective or perceptions data on governance, but rather is an issue that should be squarely addressed in all efforts to measure the quality of governance.

The full dataset of our aggregate governance indicators is available on the web at www.worldbank.org/wbi/governance/govdata/. These indicators are constructed to have a mean of zero and a standard deviation of one in each period. Actual scores range from approximately -2.5 to 2.5. In Figure 2.2 we provide a visual overview of the data for two dimensions of governance: Political Stability and Absence of Violence, and Control of Corruption. We order countries in ascending order according to their point estimates of governance in 2004 on the horizontal axis, and on the vertical axis we plot the estimate of

governance. The vertical line for each country shows the statistically likely range for the value of governance for each country, as captured by a 90% confidence interval. The size of these confidence intervals varies across countries, as different countries appear in different numbers of sources with different levels of precision. An important feature of this graph is that the confidence intervals are substantial relative to the units in which governance is measured. As a result, many of the small differences in estimates of governance across countries are not likely to be statistically significant at reasonable confidence levels. For many applications, instead of merely observing the point estimates, it is more useful to focus on the *range* of possible governance values for each country.

[figure 2.2]

In Figure 2.3 we illustrate the changes over time in our estimates of governance in individual countries, for two selected governance indicators over the period 1996-2004. In both panels, we plot the 2004 score on the horizontal axis, and the 1996 score on the vertical axis. We also plot the 45-degree line, so that countries above this line correspond to declines in the quality of governance, while countries below the line correspond to improvements in governance. Most countries are clustered quite close to the 45-degree line, indicating that changes in our estimates of governance in these countries are relatively small over the eight-year period covered by the graph. A similar pattern emerges for the other four dimensions of governance (not shown in Figure 2.3), and, not surprisingly the correlation between current and lagged estimates of governance is even higher when we consider shorter time periods.

[figure 2.3]

However, our estimates of governance do change substantially for some countries in some periods. In Figure 2.3 we have labeled those countries for which the change in estimated governance over the 1996-2004 period is sufficiently large that the 90% confidence intervals for governance in the two periods do not overlap. Although not a formal test of statistical significance, we will show later in the paper that this is a useful rule of thumb for identifying statistically and practically important changes in governance. For example, from 1996 to 2004, countries like Cote d'Ivoire, Zimbabwe, Nepal, and the Central African Republic show substantial declines in the Voice and Accountability measure, among others, while countries like Argentina and Sierra Leone deteriorate on Regulatory Quality, and Zimbabwe, Cyprus, Israel, and Moldova decline on Control of Corruption. Compare this with countries like Latvia and Bahrain that show substantial improvements in Control of Corruption, and Croatia, Nigeria, and Bosnia and Herzegovina that improved in Voice and Accountability. [3]

In the working paper version of this chapter (Kaufmann, Kraay, and Mastruzzi 2005), we investigated in more detail the factors underlying the changes in our estimates of governance. We find that for large changes in governance in either direction, there is a reassuringly high degree of consensus among our underlying data sources for each country as to the direction of the change. For a typical large change in governance, over 80 percent of the data sources available for that country move in the same direction as the aggregate indicator. Moreover, although the number of sources for our governance indicators has increased markedly over time, we show that this addition of new sources does not appear to have very substantial effects on the changes over time in the governance estimates. Taken together, this evidence suggests that for the large changes

in governance shown in this table, we can have a good deal of confidence that it is mostly driven by changes in the underlying sources on which the aggregate indicators are based. In contrast, we should be much more cautious in our interpretation of many of the smaller changes in our aggregate governance indicators.

It is important to note that our aggregate indicators are measured in relative units, since we have scaled them to have a mean of zero in each period. This opens the possibility that although many countries do not display large changes over time in their relative positions, it may be the case that there are broad-based improvements in global averages of governance that are not being picked up by our indicators. In order to determine how important this concern is, we have gone back to our underlying data sources and selected a subset of them for which we can track over time a similar specific concept of governance for a common set of countries.

In Table 2.1 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 1996-2004. Only one of them, GCS, is a survey with sufficiently standard format to enable comparisons over this period of time. The first five columns present the average across all countries of each of the sources in each of the years. The underlying data have been rescaled to run from zero to one, and for each source and governance component, we report the score on the same question or average of questions that we use in the aggregate indicator. The next five columns report the standard deviation across countries for each source. The final column reports the t-statistic associated with a test of the null hypothesis that the world average score is the same in 1996 as in 2004.

[table 2.1]

The picture that emerges from Table 2.1 is sobering. There is very little evidence of statistically significant improvements in governance worldwide. The 22 eight-year changes reported here are divided exactly in half into 11 improvements and 11 declines in global averages. There are nine 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 three improvements and six declines. It is not clear how much importance ought to be ascribed to these trends in world averages. On the one hand, these statistics represent the only information we have on trends over time, and so they should be taken seriously. On the other hand, it is 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 do not have evidence of any significant improvement in governance worldwide and, if anything, the evidence is suggestive of a deterioration in key dimensions such as regulatory quality, rule of law and control of corruption.

2. Interpreting differences in governance across countries and over time

In our description of the data in the previous section we emphasized the importance of measurement error in our governance indicators. In this section we first use the specific example of the Control of Corruption eligibility criterion for the United States Millennium Challenge Account to illustrate the importance of margins of error for cross-country comparisons of governance indicators. We also show how the presence of

margins of error affects the conclusions we can draw about the statistical and practical importance of observed changes over time in governance.

2.1. Cross-country governance comparisons and the MCA

As an illustration of the importance of margins of error in governance comparisons, consider the eligibility criteria for the U.S. Millennium Challenge Account (MCA). Countries' eligibility for grants from the MCA is determined by their relative positions on 16 different measures of country performance. One of these is our Control of Corruption indicator, where countries are required to score above the median among all potentially eligible countries in order to qualify for MCA funding. As we have noted elsewhere, this procedure risks misclassifying countries around the median because the margins of error for such countries often includes the median score. In contrast, for countries near the top and the bottom of potential MCA beneficiaries, we can be quite confident that they do in fact fall above and below the median, respectively.

Table 2.2 illustrates the role of margins of error in this calculation. We focus attention on the set of 70 countries identified as potential MCA beneficiaries for the 2005 fiscal year.[4] For these countries, we calculate the median score on our Control of Corruption indicator for 2004. Next, using our governance estimates and their accompanying standard errors, for each country we calculate the probability that the country's level of corruption falls above the median for this group. The results of this calculation are summarized in the first column of Table 2.2. For 17 poorly-performing countries, about one-quarter of the sample, there is less than a 10 percent chance that

corruption in these countries actually falls above the median. For another 23 countries, or about a third of the sample, we are quite confident that corruption in these countries falls above the median, with a probability of at least 90 percent. In contrast, for the remaining 30 countries, the probability that they fall above the median is somewhere between 10 percent and 90 percent, and so we have less confidence that these countries are correctly classified. If we relax our standards of significance to 25 percent and 75 percent, we find that only about 20 countries out of 70, or 29 percent of countries fall in this zone of uncertainty.[5]

[table 2.2]

This example illustrates the importance of taking margins of error into account when making governance comparisons across countries. Our aggregate governance indicator is able to identify with a fairly substantial degree of confidence groups of countries where the probability that corruption is above or below the median is large. But at the same time there remains an intermediate group of countries where we can be less confident that they are correctly classified as being "good" or "bad" performers based on their point estimates of governance alone.

It is also important to note how this example illustrates the benefit of aggregating many sources of data on corruption. The remaining columns of Table 2.2 perform the same calculations, but relying on successively less precise measures of governance. The second and third columns use our own Control of Corruption indicators for 2000 and 1996. These indicators cover fewer countries, and because they rely on a smaller set of sources available at the time, the margins of error for individual countries are higher than in 2004 (see the standard errors reported in the last row). In 1996, for example, 35

percent of the countries for which data is available fall in the intermediate category where the probability that they fall in the top half of the sample is between 25 percent and 75 percent – as opposed to only 29 percent of countries falling in this grey area with the 2004 indicator. The last three columns of the table show the same information for three of our individual sources, WMO, DRI, and GCS. These individual sources have substantially higher margins of error than our aggregate indicators, and in the case of DRI and GCS, also cover substantially fewer countries. In addition, we see that there is greater uncertainty about country rankings when relying on just a single indicator—for GCS, for example, the fraction of countries falling in the intermediate category rises to 40 percent. This illustrates the benefit of relying on aggregate indicators, which are more informative than individual indicators when trying to classify countries according to their levels of governance.

2.2. Margins of error and changes over time in governance

It is useful to begin our discussion with the simplest possible example of how measurement error impacts our interpretation of changes over time in observed governance indicators, both subjective and objective. Suppose that we have only one source of governance data observed at two points in time, and we want to make inferences about how governance has changed in a country. To keep notation as simple as possible, we suppress country subscripts and write the observed data at time t, $y(t)$, as the sum of true unobserved governance in that period, $g(t)$, and an error term capturing measurement error:

(1) $$y(t) = g(t) + \varepsilon(t) \quad , \quad t = 1, 2$$

As a choice of units, we assume that true governance has mean zero and standard deviation one, and that the error term has zero mean. For simplicity we assume that the variance of the error term is the same in both periods and is equal to σ^2. Note that σ^2 is the noise-to-signal ratio in the observed governance data (the ratio of the variance of the error to the variance of unobserved governance). We also allow for the possibility that both governance and the error term are correlated over time, with correlations ρ and r, respectively. Finally we assume that both governance and the error term are normally distributed. With these simplifying assumptions, consider the problem of making inferences about the change in unobserved governance, g(t)-g(t-1), conditional on observing data y(t) and y(t-1) in the two periods. Using the fact that unobserved governance and the data are jointly normally distributed, we can use the properties of the multivariate normal distribution to arrive at the following expressions for the mean and variance of the change in governance, conditional on the observed data:[6]

(2)
$$E[g(t) - g(t-1) \mid y(t), y(t-1)] = \frac{(1-\rho)\cdot(y(t) - y(t-1))}{1 + \sigma^2 \cdot (1-r) - \rho}$$

$$V[g(t) - g(t-1) \mid y(t), y(t-1)] = \frac{2\cdot(1-\rho)\cdot(1-r)\cdot\sigma^2}{1 + \sigma^2 \cdot (1-r) - \rho}$$

It is natural to use this conditional mean as our best estimate of the change in governance, and the conditional variance as an indicator of the confidence we have in the estimate.

This is in fact exactly analogous to how we obtain estimates of levels of governance and associated standard errors using the unobserved components model.

To interpret these expressions, consider first the case where there is no persistence in governance or in the error terms, i.e. $\rho=r=0$. In this case, our estimate of the change in governance is simply $\frac{y(t)-y(t-1)}{1+\sigma^2}$. In particular, we should take the observed change in the single source and scale it down by a factor of $\frac{1}{1+\sigma^2}$ to reflect the fact that the data measures governance with error. It is also clear from equation (2) that the higher is ρ, the more we should discount observed changes in governance. Intuitively, if we knew that governance changes very slowly over time, then any observed change in the data is more likely to reflect changes in the error term, and so we should discount this observed change more heavily. In the limit where governance is perfectly correlated in the two periods, we would know for sure that any change observed in the data must reflect only fluctuations in the error term, and so we would completely discount the observed change in the data. That is, our estimate of the change in governance would be zero regardless of the observed change in the data.

The effect of persistence in the error terms works in the opposite direction: we should scale down the observed change in the data by less the larger is the correlation over time in the error terms. Again the intuition for this is simple – if we know that the error with which a given source measures governance is persistent over time, then any observed change in the source is likely to understate the true change in unobserved governance. As a result our best estimate of the change in governance will be larger than the observed change in the data. Interestingly, if the correlation in unobserved

governance and the error term are equal to each other, i.e. ρ=r, then these two effects offset exactly and the discount applied to the observed change in governance is $\frac{1}{1+\sigma^2}$.

How much confidence should we have in the statistical significance of the change in unobserved governance based on the observed data? Suppose that we observe a change in the indicator equal to k standard deviations of the changes in this variable, i.e. $y(t)-y(t-1)=k\cdot\sqrt{2\cdot\left(1+\sigma^2\cdot(1-r)-\rho\right)}$. Does this signal a significant change in governance? In order to test the null hypothesis that the change in governance is zero, we can construct the usual z-statistic associated with this hypothesis, i.e. the ratio of the mean of the change in governance conditional on the data to the square root of the conditional variance, which simplifies to:

$$(3)\qquad z=\frac{E\left[g(t)-g(t-1)\,|\,y(t),y(t-1)\right]}{\sqrt{V\left[g(t)-g(t-1)\,|\,y(t),y(t-1)\right]}}=\frac{k}{\sigma}\cdot\sqrt{\frac{1-\rho}{1-r}}$$

Not surprisingly, the observed change in the data is more likely to signal a significant change in unobserved governance the larger is the observed change in the data (i.e. the larger is k), and the lower is the signal-to-noise ratio in the data (i.e. the smaller is σ). And building on the intuitions above, the observed change in the data is also more likely to signal a significant change in unobserved governance the lower is the persistence in unobserved governance, ρ, and the higher is the persistence in the error term, r.

Figure 2.4 puts some numbers to this simple calculation. We graph the number of standard deviations of the observed change in the data, k, on the horizontal axis, and we plot the z-statistic in Equation (3) on the vertical axis for different values of the key

parameters. We set σ^2=0.36, as this is the median value for the noise-to-signal ratio across all of the individual data sources we use to construct our six governance indicators in each of the five periods. In an earlier paper we argued that the noise-to-signal ratio in objective measures of governance is likely to be at least as large as this.[7] The thin upward-sloping line traces out the z-statistic as a function of k for this value of the noise-to-signal ratio, but assuming that the correlation in governance and the error term are zero, i.e. ρ=r=0. The z-statistic is greater than the 90-percent critical value for changes in the observed data that are more than one standard deviation away from the mean change. This suggests that if there is no persistence in governance or in the error terms, quite a large proportion of observed changes in individual governance indicators would in fact signal a significant change in unobserved governance. In fact, if changes in the observed governance indicator are approximately normally distributed, the largest one-third of all absolute changes would signal changes in governance that are significant at the 90% level.

[figure 2.4]

The bold upward-sloping line corresponds to the more empirically relevant case where there is persistence in both governance and the error terms. The line is drawn for the same noise-to-signal ratio as before, and in addition we assume that the correlation of unobserved governance over time is ρ=0.9 and the correlation in the error term is r=0.4. In the next subsection we show how these parameters can be estimated using our governance data, and find that these values are typical ones. In particular, we shall see shortly that unobserved governance tends to be highly persistent over the eight-year period spanned by our dataset, and although the error terms are also typically positively

correlated over time they are much less so than governance. Based on the intuitions developed above, this suggests that much larger observed changes in governance indicators would be required to signal statistically significant changes in unobserved governance. This is exactly what we find. The bold line crosses the 90% critical value at k=2.5, indicating that only those observed changes in the data more than 2.5 standard deviations away from the mean would signal a statistically significant change in governance. Again, if changes in the observed governance indicators are normally distributed, this would imply that only the top one percent of all absolute changes would correspond to significant changes in governance. This in turn suggests that drawing conclusions about changes in governance based on changes in individual governance indicators should be done with an abundance of caution.

In the appendix to this chapter we extend the discussion above to the case of aggregate governance indicators. The basic insights from this discussion of changes in individual indicators also carry over to changes in aggregate governance indicators. Just as we found that aggregate indicators are more informative about levels of governance than individual indicators, so changes over time in aggregate indicators can be more informative about trends in governance than changes in individual indicators. And as suggested in the discussion above, there is a tension between persistence in governance and persistence in measurement error in the aggregate indicators. The greater is the former, the more cautious we should be about observed changes in governance. And the greater is the latter, the more likely it is that observed changes in indicators of governance signal significant changes in true governance. As shown in the appendix, we find that that the simple rule of thumb we proposed above -- that changes in governance are

significant if the 90 percent (or 75 percent) confidence intervals in the two periods do not overlap -- does a fairly good job of identifying changes that are statistically significant using more formal criteria.

3. Subjective and objective measures of governance

In this section we address a number of issues that arise in using subjective or perceptions-based data to measure governance across countries. We begin by discussing why subjective data is often either the only type of data available to measure governance or else adds valuable insights over available objective measures. We next emphasize that margins of error are not unique to the subjective measures of governance that we construct, but are pervasive in all efforts to measure governance. We present some simple calculations which show that margins of error in objective measures of governance are comparable to those we present for our subjective measures. We then turn to a deeper investigation of one source of discrepancy between subjective and objective indicators, which is that the latter tend to emphasize de jure rules on the books while the former tend to pick up the de facto reality on the ground. We finally briefly describe an earlier effort of ours to quantify the importance of ideological biases in subjective measures of governance in which we found that they were small.

3.1. Perceptions matter

In this subsection we discuss some of the advantages of the subjective or perceptions-based measures of governance we use to construct our aggregate governance indicators. The primary reason for choosing subjective measures is that for many of the key dimensions of governance, such as corruption or confidence in the protection of property rights, relevant objective data are almost by definition impossible to obtain.

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

this idea are not available – nor are they likely to be, given the major data requirements for this kind of exercise.

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

More generally, subjective perceptions of governance often matter as much as the legal reality. For example, on the basis of firms' perceptions of the undue influence of powerful firms on the political decision-making process—influencing laws, policies and regulations—Hellman and Kaufmann (2003) develop a measure for 'crony bias,' or unequal influence across firms. The authors find a consistent pattern in which perceived unequal influence has strongly negative impact on the firm's assessment of public institutions, which in turn affects the behavior of the firm towards those institutions. Crony bias at both the firm and the country level is associated with lesser use of the

courts by the firms to resolve business disputes, lower enforceability of court decisions, lower levels of tax compliance, and higher levels of bribery. Thus, the evidence suggests that the inequality of influence not only damages the credibility of institutions among less (politically) powerful firms, but also affects the likelihood that they will use and provide tax resources to support such institutions, thereby perpetuating the weakness of such institutions and likelihood of capture by influential private actors.

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

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

3.2. Margins of error are not unique to subjective data

We have argued that one of the strengths of the governance indicators reported in this paper is that we are able to construct explicit margins of error associated with our

estimates of governance for each country. However it is worth emphasizing that these margins of error are not unique to subjective or perceptions-based measures of governance, but are also present -- if not explicitly noted -- in most other measures of institutional quality, or in any other socioeconomic indicator for that matter. One need only consider the range of "preliminary" estimates of basic objective variables such as real GDP growth produced in industrial countries with high-quality statistical systems to realize that measurement error in objective data is in fact pervasive and should be taken seriously.[8]

Consider for example the recent interest in constructing objective measures of governance that do not exclusively rely on perceptions-based data sources as we do, but rather on objective and quantifiable data. Several of these are described in Knack and Kugler (2002). They argue that variables such as the waiting time required to obtain a telephone line, and the number of telephone faults can serve as proxies for public administrative capacity. The reliance of the government on trade taxes can serve as a proxy for the (in)ability of the government to broaden its tax base. The volatility in budgetary expenditure shares, and similarly, the volatility of revenue shares, is indicative of a volatile and unpredictable policy environment. They also draw on a number of other measures of institutional quality pre-existing in the literature. Clague, Keefer, Knack and Olson (1996) argue that the fraction of currency in circulation that is held in the banking system is a good proxy of the extent to which individuals in a country can be confident that their property rights are protected. Finally, in a series of papers, Djankov et al (2002, 2003) compile cross-country data on the number of administrative procedures required to

start a business, and the number of legal procedures required to collect an unpaid debt. These measures capture the complexity of the regulatory and legal environment.

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

This is of course not surprising, nor should it be considered a drawback of such measures -- all of which are necessarily imperfect proxies for broader notions of governance. However, it does mean that one should consider seriously the margins of error for objective indicators as well, to the extent that these are used as proxies for broad concepts of governance such as the ones we measure using subjective data in this paper.[9] Although these margins of error are generally not made explicit for objective indicators, a simple calculation can give a sense of their order of magnitude. Suppose that we have two noisy indicators y on a common unobserved concept of governance, g, i.e.: $y_i = g + \varepsilon_i$, i=1,2. Then if we normalize the variance of the unobserved measure of governance to be one, the correlation between the two observed indicators will be $\rho = \left(\left(1+\sigma_1^2\right)\cdot\left(1+\sigma_2^2\right)\right)^{-1/2}$. Suppose that indicator 1 is one of our subjective governance

indicators, for which the variance of the measurement error, σ_1^2, is known, and that indicator 2 is one of the objective indicators described above. Then from the observed correlation between the two indicators, we can infer the variance of measurement error in the objective indicator, σ_2^2.

The results of this calculation can be found in Table 2.3. The rows of Table 2.3 correspond to the various objective governance indicators discussed above. In the first two columns, we identify the objective indicator, and the subjective aggregate governance indicator which best corresponds to it. In the third column we report the correlation between the subjective and the objective indicator, using our 2002 governance indicators. The next three columns report the implied standard deviation of measurement error in the objective indicator, under three assumptions: (A) that our estimate of the standard deviation of measurement error in the subjective indicator is correct, (B) that the subjective and objective indicators have the same standard deviation of measurement error, and (C) that the standard deviation of measurement error in the subjective indicator is twice as large as that in the objective indicator. Finally in the last column we report the actual standard deviation of measurement error, computed as the average across all countries of the country-specific standard errors in our governance indicators.

[table 2.3]

The results in Table 2.3 are quite striking. For all indicators, and for all three sets of assumptions, the implied standard deviation of measurement error in the objective indicators is very high relative to the corresponding standard deviation of the subjective governance indicators. Under the benchmark assumption (A) which takes seriously the margins of error we have computed for our governance indicators, we find that the

implied margin of error for the objective indicators is between seven and 15 times larger than that of the subjective indicators. This clearly exaggerates the difference in precision between subjective and objective indicators because we are comparing a single objective indicator with an aggregate of several subjective measures and, as discussed earlier, we should expect aggregation to improve precision. But this is only part of the story. For the GE and RQ indicators, we have a median of six sources per country, while for RL we have a median of eight sources. This can explain why the standard deviation of measurement error of the objective sources might be $\sqrt{6} = 2.4$ to $\sqrt{8} = 2.8$ times higher than that of the corresponding subjective indicators, but still cannot explain all of the difference in the precision of the indicators that we see. Similarly, the last row in Table 2.3 reports the correlation of GE with an aggregate of all the objective indicators. In this case, the benefits of aggregation would be roughly comparable for the two indicators, with a median of five sources per country for the objective indicator and a median of six sources per country for GE. Nevertheless, we find that the implied standard deviation of measurement error is still four times as large for the objective indicator as it is for the subjective one.

Assumptions (B) and (C) are designed to be more favorable to the precision of the objective indicators. Assumption (B) discards the information in the margins of error that we have constructed for the subjective indicator, and simply makes the neutral assumption that the subjective and the objective indicators have the same standard deviation of measurement error. This reduces the implied standard deviation of measurement error for the objective indicator relative to the benchmark assumption (A), but it remains large at 0.6 for the composite objective indicator, and higher for the

individual indicators. Assumption (C) weights things even further in favor of the objective indicators, assuming that the objective indicator is twice as precise as the subjective indicator. In this case, we continue to find very substantial estimates of the standard deviation of measurement error, on the order of 0.4 and higher for individual objective indicators.

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

3.3. *De jure* and *de facto* governance indicators

A recurrent theme in this paper is that individual sources of governance data are imperfect and provide only noisy signals of unobserved governance. In the previous subsection we saw that part of this measurement error is due to the fact that all specific subjective and objective measures of governance are imperfect proxies for the broader concepts of governance that they are used to measure. In this section we turn to a different source of measurement error arising from the distinction between *de jure* and *de facto* measures of governance. Consider for example the very useful "Doing Business" project of the World Bank, which has compiled objective measures of various dimensions of the regulatory environment across countries by interviewing law firms around the world about formal rules and regulations in their countries. As with the subjective measures of ease of business entry, there are gaps between this specific dimension of

regulation and the overall quality of the regulatory environment. Interestingly, as we are about to see, there are systematic differences between even very specific subjective and objective measures, which reflect the sometimes wide gap between the *de jure* rules on the books and their *de facto* application.

We consider two measures of the *de facto* environment facing firms taken from the survey of over 8,000 firms in 104 countries carried out by the World Economic Forum in 2004 as an input to their Global Competitive Report (GCS). These two variables capture firms' assessments of the ease of starting a business, as well as their reported tax burden.[10] We then match these with two closely-related *de jure* measures from other sources. For ease of starting a business, we draw on the Doing Business (DB) project at the World Bank discussed above. From this dataset we take the number of days required to start a business. For perceptions of the tax burden, we have independently collected statutory tax rates by sector and size of firm for each of the countries in the sample. We then assign these tax rates to individual firms, and then average these up to the country level to obtain average measures of the statutory tax burden.[11] Figure 2.5 plots the measure against the number of days to start a business from the DB dataset.

[figure 2.5]

We begin the statistical analysis with simple ordinary least squares regressions of perceptions of ease of starting a business on the corresponding objective measure (first column of Table 2.4). Not surprisingly, the objective measure enters negatively and is highly statistically significant with a t-statistic of more than five, indicating that firms perceive it more difficult to start a business in countries where the number of days

required to do so is large. More interesting for our purposes is the observation that the R-squared of the regression is very modest, at only 0.23. We cannot say at this point whether this results reflects measurement error in the subjective or the objective measure. One hypothesis, however, is that the objective measure fails to capture the extent to which the formal requirements to start a business are altered by the presence of corruption or other forms of informality in their application. To investigate this possibility we add our aggregate measure of Control of Corruption to the regression.[12] We find that this variable enters positively and highly significantly, indicating that perceptions of the ease of starting a business are significantly better in countries with less corruption, even after controlling for the *de jure* rules governing business entry. Once we add corruption, the coefficient on the *de jure* rules falls by half, and its significance also drops to the 10 percent level. Moreover the adjusted R-squared of the regression doubles to 0.44, indicating substantial explanatory power for this additional variable.

[table 2.4]

There is, however, an obvious difficulty with this result. It could well be the case that firms' responses to the question regarding business entry are non-specific, in the sense that they will provide low responses if their assessment of the overall business environment is negative. This generalized dissatisfaction could account for the significance of the corruption variable, rather than any effect of corruption on the enforcement of business entry procedures. We address this possibility in the next three columns. One test for this problem of non-specificity is to ask whether unrelated objective measures of the business environment also predict perceptions about ease of entry. We do this in the third column by adding the objective tax burden question to the regression. If firm responses reflect generalized dissatisfaction, we might expect this

variable also to enter significantly, yet it does not. In the fourth column we instead add firms' responses to a question about the overall regulatory environment that they face. Again we find that corruption remains highly significant, and in this case the general question about regulation is also highly significant. This suggests that while non-specificity of responses may be a concern, it does not fully account for the significance of the corruption measure in the previous specifications. Interestingly, in both specifications, we find that the coefficient on the objective entry measure becomes larger and more significant as we add these control variables. Finally, we note that all these results go through when we put all four variables in the regression.

The second and third panels of Table 2.4 reveal interesting differences between developing countries on the one hand, and OECD and newly-industrialized countries on the other. In the developing country sample, the results described above go through for the most part. However, it is interesting to note that the magnitude and significance of the objective measure is in general smaller in the developing country sample, and larger in the industrial country sample, while the converse is true for the corruption variable. Taken together these results suggest that firm perceptions of the ease of starting a business depend on both *de jure* rules, as well as the institutional environment in which those rules are applied. Moreover, the relative importance of *de jure* rules seems to be higher in industrial than in developing countries. More broadly, the lesson from this simple exercise is that it can be misleading to rely exclusively on either perceptions of *de facto* governance or objective measures of the *de jure* rules.

We perform the same sequence of regressions using the question on perceptions of tax burdens from GCS as the dependent variable. The results are broadly similar to

those discussed above and are reported in the continuation of Table 2.4. In the full sample of countries, we find that perceptions of tax burdens are strongly correlated with our *de jure* measure of statutory tax rates. While in the full sample of countries we do not find corruption to enter significantly, it does in the developing country sample where we might expect corruption to matter more for perceptions of the tax burden. As before, we address the possibility that the tax burden question captures generalized dissatisfaction rather than a specific concern with taxation by including the objective measure of days to start a business, and we find that the corruption variable remains significant. Also consistent with our priors, we find that differences in statutory tax rates have much stronger explanatory power for perceptions of tax burden in the industrial country sample. Although the overall results are not quite as strong as for the business entry example discussed above, the picture that emerges is qualitatively quite similar.

In sum, the results suggest that assessments of governance should not be based solely on objective measures of the *de jure* situation. We have seen that firms' perceptions of the ease of starting a business and the weight of their tax burden depend not only on the *de jure* regulations that they face, but also on the environment in which these regulations are applied. Laws and regulations are often adopted but their implementation subverted due to prevailing informal mechanisms. In these settings the essence of how policies and regulations are actually implemented may be frequently missed by objective indicators. This is not to say, of course, that firm-based surveys of perceptions are devoid of margins of error and related challenges. Rather, the results we have shown emphasize the importance of relying on a range of measures to assess governance and on recognizing that no single measure is a perfect proxy for governance.

3.4. Potential ideological biases

We conclude this section by briefly addressing the critique that subjective data from polls of experts may reflect the ideological tendencies of the institutions compiling the performance ratings. Our assumption has been that this is not a major concern for the sources on which we rely. This is because we find a very high degree of correlation among most of our sources, which is difficult to reconcile with a systematic ideological bias present in certain sources. In a previous paper (Kaufmann, Kraay, and Mastruzzi (2004)) we nevertheless took this possibility seriously and investigated the extent to which the differences in assessments across sources are related to observable measures of the ideology of the government in power in each country.

We approached the question as follows. Our identifying assumption was that surveys of firms or individuals are not tainted by ideology, since they reflect the views of a large number of respondents in each country. In contrast, it is possible that the views of a smaller number of raters affiliated with a particular institution may reflect the ideology of that group. We can therefore identify the effects of ideology by looking at the correlation across countries between the ideology of the government in power, and the *difference* in the percentile ranks assigned to countries by a poll of experts and a survey of individuals and firms. We implemented this idea using the World Bank's Business Environment Survey (WBS) for 2000 and an independently available indicator variable that takes on the value 1 if the government in power is left-of-center, 2 if it is center, and 3 if it is right-of-center.[13] The coefficient on the ideology variable was intended to

capture the extent to which a given poll of experts rates countries with left- or right-wing governments systematically differently from a survey.

The results of the regressions confirm that most sources are not affected by ideological bias. Only one source, Heritage Foundation, was found to assign relatively higher scores to countries with right-of-center governments than the corresponding surveys. However, it is worth emphasizing that this "ideology bias" is fairly modest in magnitude. The coefficient estimates indeed indicate that a country with a right-of-center government would get between 7 and 10 percentile points higher than a center government. Moreover, in all cases, the ideology variable in a statistical sense explains only a trivial fraction of the difference in assessments between polls and surveys, suggesting that the importance of ideological biases in polls is quite small overall[14].

4. Interpreting governance-income correlations

We began this chapter by noting the strong consensus that governance matters for economic development. An important part of the evidence in support of this view comes from providing a causal interpretation of the strong observed positive correlation between governance and per capita incomes across countries. But there are alternative interpretations of this correlation. We first consider – and discount – the interpretation that these strong correlations are a consequence of "halo effects," i.e. an upward bias in perceptions of governance in rich countries simply because they are rich. We also discuss – and refute – the argument that the weak governance performance of countries in Africa should be discounted in some sense because these countries are poor.

4.1. Halo effects

Perceptions-based measures of governance such as the ones we develop here are potentially subject to a number of biases. One common critique is that perceptions of governance are biased upwards in rich countries because respondents view the development success of the country in question as evidence that institutional quality is good. This type of bias is sometimes referred to as a "halo effect."[15] This in turn implies that part of the observed high correlation between per capita incomes and governance spuriously reflects this bias.

To formalize the idea of halo effects, suppose that we can write our observed estimates of governance, g*, as the sum of true governance, g, and an error term, u:

(4) $g^* = g + u$

The essence of the halo effect argument is that this error term u is correlated with per capita incomes, y. The relevant question then is the extent to which this spurious correlation can account for the high observed correlation between measured governance and per capita incomes. Intuitively, it should be clear that in order for halo effects to substantially account for the correlation between incomes and measured governance, it must be the case that the correlation between the error and income is large. Perhaps less obviously, it must also be the case that the variance of the error term is large relative to the variation in governance. Otherwise, even if the error term is strongly correlated with

income, the fact that it accounts for little of the variance in measured governance means that it will have little impact on the correlation between measured governance and per capita income. Our argument in a nutshell is that for reasonable assumptions on the importance of measurement error, this measurement error would have to be implausibly highly correlated with per capita incomes in order to constitute a significant source of bias.

To formalize this intuition, we decompose the observed correlation between measured governance and per capita income into a term reflecting the true correlation between governance and income and a term attributable to the halo effect:

$$\text{(5)} \qquad \text{CORR}(g^*, y) = \sqrt{1-s} \cdot \text{CORR}(g, y) + \sqrt{s} \cdot \text{CORR}(u, y)$$

where $s = V[u]/V[g^*]$ is a measure of how noisy the governance indicator is. Note also that the correlation between measured governance and per capita income that we see in the data is around 0.8.

To understand this expression, suppose that the true correlation between governance and income were zero, so that all of the observed correlation between income and governance is due to the second term capturing halo effects. This consists of the actual correlation of the error term with per capita income, which is multiplied by the square root of the share of the variance in governance due to the error term. Suppose that the governance indicator is very noisy so that the share of the variance approaches one. Then the correlation of the error term with per capita income must be equal to the observed correlation in the data. Suppose however that the governance indicator is at

least somewhat informative, so that s is less than one. In order to match the observed correlation in the data, the halo effect correlation in the error term must be even larger than the 0.8 observed in the data. This example illustrates how the importance of halo effects in accounting for the observed correlation between governance and per capita income depends on both the strength of the halo effect itself, as well as the relative importance of measurement error in the governance indicator.

This example is extreme because we have assumed that the true correlation between governance and income is zero. We now relax this assumption and revisit the question of how strong halo effects need to be to account for the observed correlation between measured governance and per capita income of 0.8. We do this with the help of Figure 2.6, which graphs the strength of the halo effect, i.e. CORR[u,y], on the vertical axis, against the share of the variance in governance due to the residual, i.e. s, on the horizontal axis. The different lines on the graph correspond to different assumptions for the true correlation between governance and income. We have already discussed the intuition for the case where this correlation is zero, shown as the highest line in the graph. If the share of the variance in governance due to measurement error is one, the halo effect correlation must be equal to 0.8. As we move to the left and the governance indicator becomes more informative, the required correlation increases.

[figure 2.6]

The lines corresponding to successively higher true correlations between governance and income fall everywhere below the first series. This is because once we allow for some correlation between true governance and income, the halo effect needed to account for the correlation between observed governance and income is weaker.

Interestingly, however, even if the true correlation is quite substantial at 0.6, the lowest line in Figure 2.6 tells us that halo effects must still be quite considerable, with a correlation of at least 0.5, to match the observed data.[16] This lower bound occurs for intermediate values of the share of the variance of governance due to measurement error. It is also interesting to ask what a reasonable value for this share might be, in order to pin down more precisely how strong halo effects must be. One way to do so is to consider the standard errors of the governance estimates, which average around 0.25 as compared with the standard deviation of measured governance of 1. This suggests that the share of the variance of governance due to the error term is in fact quite small at $s=0.25^2=0.06$. For this low variance share, the halo effect correlation would need to be 0.9 in order to match the observed data. If the true correlation between governance and income were much lower, for example at 0.4, then even if measurement error in governance were perfectly correlated with per capita income it would not be possible to generate the observed correlation between governance and per capita incomes.

This strong conclusion is driven by the assumption that that measurement error accounts for a relatively small portion of the variation in observed governance. As a result this measurement error needs to be very highly correlated with incomes in order to match the data. One could argue that we are understating the importance of measurement error by relying on the estimated standard errors from our governance indicators. After all, these are based on the assumption that measurement error is uncorrelated across different sources of governance data. However, if halo effects are important, the measurement error in individual sources will be correlated not only with per capita income, but also with other sources.. This in turn would imply a greater imprecision of

the governance estimates. To capture this possibility, suppose that the standard error of the governance estimates were twice as large as what we actually have, at 0.5.[17] This implies s=0.25, and for this value of s we can see from Figure 2.6 that the halo effect correlation would still need to be very high at almost 0.6 in order to match the data.

In summary, these results suggest to us that although halo effects may well be present in perceptions-based measures of governance, these halo effects need to be implausibly strong in order to impart a substantial upward bias in the correlation between measured governance and per capita incomes. Moreover, it is worth noting that there may well be other factors offsetting such halo effects. One is the tendency of survey respondents in developed countries to be particularly critical of their own institutions.[18] It is also worth noting that some cross-country polls of experts deliberately apply higher standards to rich countries when assessing their governance.[19] Overall, then, we do not think that halo effects are a significant source of bias in the correlations between governance and per capita incomes our data.[20]

4.2. Controlling for income in governance comparisons

In a recent paper, Sachs and others (2004) have argued that weak governance is not a major factor in Africa's poor growth performance. The argument is that, once we control for per capita income, countries in Sub-Saharan Africa do not have particularly poor governance indicators. This point is illustrated in Figure 2.7, which plots our 2004 Rule of Law measure (on the vertical axis) against the logarithm of real per capita GDP in the mid-1990s (on the horizontal axis). Note that the per capita income variable has been

rescaled to have mean zero and standard deviation of one, as the governance indicator has. A striking observation from this graph is that over half (27 out of 46) of the countries in usb-saharan Africa actually fall above the simple ordinary least squares regression line. At first glance, this appears to lend credence to the argument that governance in Africa is on average what one might expect given the region's low income levels.

[figure 2.7]

However, it is misleading to conclude from this simple graph that Africa's governance performance is reasonable given its per capita income. This interpretation of the graph is valid only to the extent that the OLS regression line would capture a causal relationship from higher income to better governance. But a large body of research indicates that there is substantial causation in the other direction as well – better governance leads to higher incomes. Moreover, the magnitude of the estimated effect of governance on per capita incomes in the long run is large.[21] Available estimates suggest that a one standard deviation improvement in governance would lead to a two- to three-fold difference in income levels in the long run. A one standard deviation change in governance would correspond to, for example the difference between Kenya and Turkey on our 2004 Rule of Law indicator. This means that the simple OLS relationship will exaggerate the positive effects of income on governance because it also reflects the strong effect in the opposite direction, from governance to incomes. In order to compare governance in Sub-Saharan Africa to what might be expected given income levels, we need to first isolate these two directions of causation, so to be able to focus in particular on the causal effect of income on governance.

The lines in Figure 2.7 show two alternative estimates of the causal effect of income on governance. The (slightly) upward-sloping one comes from Rigobon and Rodrik (2004). They study the causal relationships between per capita income, democracy, rule of law, openness to international trade, and geography, using identification through heteroskedasticity to isolate the causal effects.[22] As expected, this line is substantially flatter than the OLS regression line, consistent with the intuition that the latter relationship overstated the true causal effect of incomes on governance. This flattening has important consequences for our conclusions about the quality of governance in Africa controlling for income levels. Once we isolate this much weaker effect of income on governance, we find that only seven out of 46 countries in the region fall above the regression line: Ghana, Lesotho, Cape Verde, Namibia, South Africa, Botswana, and Mauritius. In contrast, the vast majority of countries in Africa have governance that is worse than their income levels would predict.

The weakly downward-sloping line presents another estimate of the effect of income on governance, coming from Kaufmann and Kraay (2002). In that paper we used a different approach to identification and found a zero or even negative impact of income on governance. Although this finding may be somewhat extreme, it leads to the same conclusions regarding the quality of governance in Africa – now only six out of 46 countries in the region fall above the regression line, indicating governance levels better than what per capita incomes would predict.

Overall this evidence suggests that it would be inappropriate to discount the governance performance of countries in Sub-Saharan Africa based on their low income levels. The reason is simple. The only way to justify such a discount is to argue that

higher incomes exert a positive causal effect on governance. But available evidence suggests that the causal impact of incomes on governance is small. Rather, the observed correlation between governance and per capita incomes primarily reflects causation in the other direction: better governance raises per capita incomes.

Conclusions

There is by now broad consensus among academics and policymakers alike that good governance matters for economic development. There is also growing awareness in the aid community that good governance matters for the effectiveness of development assistance. In light of these facts it is important to be able to measure levels and changes over time in governance across countries. This paper summarizes our recent work to construct aggregate governance indicators which seek to provide such information. Relative to previous years, these indicators reflect a significant expansion of our underlying data set of several hundred individual variables measuring perceptions of governance, drawn from 37 separate data sources.

In our work we have emphasized the difficulty of measuring governance. We have argued that one of the strengths of our composite governance indicators is that they can be more informative than individual data sources—on average the aggregation reduces the margin of error by about one-half. Further, given the increasing number of separate data sources now at our disposal to construct these aggregate indicators, we find that the margins of error of the latest period under measure are smaller than in earlier periods. However, these margins of error, even in our most recent aggregate indicators,

still remain substantial, and thus all our previous cautionary suggestions regarding interpretation continue to apply.

At the same time, we have emphasized that these margins of error are not unique to perceptions-based measures of governance, but are an important feature of all efforts to measure governance. In fact, we have argued that, for the purposes of measuring governance, there are few alternatives to the subjective, experiential data on which we rely. Even in cases where objective indicators of governance are available, we have noted that these too have implicit margins of error, and we have provided calculations indicating that these margins of error are on the same order of magnitude as those associated with our subjective aggregates. We have also provided evidence that the type of perceptions data on which we rely provides insights into governance that are difficult to obtain from more objective or quantifiable measures. For example, we have shown that firms' perceptions of the difficulty of starting a new business, or of their tax burdens, do not depend solely on the relevant legal framework governing business entry and taxation, but are also influenced by the degree of corruption in their country. This suggests that not only formal rules matter, but also the institutional environment in which these rules are applied and enforced. Thus, wherever objective data on governance or investment climate are collected, a comprehensive analysis of governance and institutional change ought to be complemented by data from the reports of the economic agents on the ground, such as firms or users of services, which inevitably will contain an element of subjectivity. Finally, to corroborate the relevance and validity of using subjective data, we have also empirically investigated, and for the most part discounted,

the importance of ideological biases in the perceptions data from polls of experts on which we rely.

Policymakers are often particularly interested in *trends* in institutional quality: is governance improving or worsening over time in a particular country? As we have emphasized in our work, the presence of measurement error in all types of governance indicators, including our own, makes assessing trends in governance a challenging undertaking. In this paper we developed a formal statistical methodology, as well as a simple rule of thumb, for identifying changes in governance that are likely to be statistically and practically significant. Over the eight-year period from 1996-2004 spanned by our governance indicators, we find that in about 5 to 7 percent of countries we can be confident (at the 90 percent significance level) that governance has changed substantially. And at a lower 75 percent significance level, roughly 20 percent of all observed changes stand out as significant. Importantly, we show that there is a great deal of agreement among our many data sources about the direction of change in governance in these countries. Overall this reminds us that while change in institutional quality often takes place haltingly, gradually, or not at all, there are also countries where one can point to sharp improvements or deteriorations even over a fairly short eight-year period. Significant and rapid institutional change, while not the norm, is feasible and does take place in practice.

Finally, we have discussed two important issues that arise in interpreting the strong positive correlation between governance and income levels. Some observers have argued that these positive correlations are substantially due to "halo effects" – perceptions of governance in rich countries are good simply because the countries are

rich. We have argued that such halo effects would need to be implausibly large to account for cross-country correlations between governance and incomes.

We have also considered the frequently-heard argument that poor levels of governance should be significantly discounted where the country is poor. Put differently, to what extent does it make sense to ask whether a country is well or poorly governed *given* its income level? This issue is often raised in the context of Sub-Saharan Africa, where too many countries are both very poor and very poorly governed. We make the simple observation that in order to answer this question, it is necessary to isolate the causal impact of income levels on governance. Simply relying on the observed correlation is inappropriate, as much of this reflects strong causal effects running from governance to per capita incomes. While identifying the effects of income on governance is difficult, the few available estimates suggest that this feedback effect is minimal. As a result, there is little basis on which to argue that the poor governance performance of many countries in Sub-Saharan Africa should be discounted simply based on low income levels.

In conclusion, it is important to keep some perspective on this contribution. While these aggregate governance indicators have been useful for eight years in providing a general snapshot of the countries of the world for various broad components of governance, and while their margins of error have declined over time, they remain a rather blunt instrument for specific policy advice at the country level. As we have argued in the past, these aggregate indicators need to be complemented with in-depth in-country governance diagnostics, based on micro-surveys of households, firms, and public officials within the country. The lessons learned from these combined aggregate and

micro-data sets do point to the importance of moving concretely to the next stage of governance reforms in Africa and elsewhere. These lessons are, among others, to stress reforms to increase transparency (such as natural resource revenue transparency mechanisms, disclosure of assets of politicians, voting records of parliamentarians, political campaign contributions, and fiscal accounts), to alter incentives in institutions so as to increasingly focus on prevention and deterrence (rather than overly relying on prosecutions), and to work more closely with other key actors outside the public sector as well, such as the heretofore neglected private sector.

NOTES

[*] This chapter is substantially based on our earlier paper "Governance Matters IV: Governance Indicators 1996-2004" available at http://www.worldbank.org/wbi/governance/pubs/govmatters4.html. We refer the interested reader to this paper for more details on our governance indicators and the underlying data. We would like to thank S. Radelet for excellent feedback, and M. Levy, G. Dunn, A. Karatnycky, R. Fullenbaum, A. Williamson, A. Bellver, S. Weber, D. Cingranelli, D. Richards, R. Writer, M. Wolkers, C. McLiesh, M. Gibney, C. MacCormac, M. Seligson, E. Kite, E. Hart, T. Sealy, D. West, M. Carballo, F. Ndukwe, 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. The views expressed here are the authors' and do not necessarily reflect those of the World Bank, its Executive Directors, or the countries they represent.

[1] These are Acemoglu, Johnson and Robinson (2000), Kaufmann and Kraay (2002), Alcala and Ciccone (2004), and Rodrik, Subramanian and Trebbi (2004).

[2] This is just one example. Others include the use of the World Bank's Country Policy and Institutional Assessment (CPIA) ratings to determine the allocation of highly-concessional lending in low-income countries, and the use of our indicators by the Netherlands development agency for monitoring governance in countries where it is active.

[3] Focusing on the shorter 1998-2004 period (which has a larger country overlap) also yields a number of countries that have undergone large changes, such as the decline

exhibited in Control of Corruption, Government Effectiveness, and Rule of Law for West Bank/Gaza, Ivory Coast, Zimbabwe, and Eritrea, and the deterioration in Voice and Accountability in Nepal, Kyrgyz Republic, and Russia. The Slovak Republic, Croatia, Serbia, Bulgaria, Madagascar and Colombia all realized improvements over this period in Control of Corruption, as did Rwanda, Sierra Leone, Angola, Turkey, South Africa and Senegal in Political Stability/Violence.

[4] See http://www.mcc.gov/ for details on the MCA eligibility criteria.

[5] We first performed these MCA-related calculations in late 2002, shortly after the announcement of the initial MCA eligibility criteria. At that time, using the older version of our 2000 Control of Corruption indicator, we found that 23 out of 61 countries (or 38 percent of countries) fell in this intermediate zone. This much higher proportion of intermediate countries reflected the fact that the old version of or 2000 Control of Corruption indicator relied on substantially fewer data sources than we now have available to us for both 2000 and 2004.

[6] The simple example here is a special case of a more general model we discuss below.

[7] See Kaufmann, Kraay and Mastruzzi (2004)

[8] The discussion in this subsection is taken from Kaufmann, Kraay, and Mastruzzi (2004). The calculations involving the governance indicators here are based on the 2002 indicators that were the latest available at that time.

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

[10] For the past number of years, collaboration between WBI and the WEF has resulted in an in-depth coverage of governance in the survey, and in the WBI contribution of a governance chapter for each GCR. For details on the data we use for the text described above, and the related coverage of these governance issues at the micro-level, see the Governance chapter in the GCR 2004, at http://www.worldbank.org/wbi/governance/pubs/gcr2004.html.

[11] The main source for the effective tax rates was the PricewaterhouseCoopers report "Corporate taxes: worldwide summaries (2003-2004)", covering 85 of our sample of 104 countries. As some countries have differential tax rates, to map the country-level data from the report to the individual firm-level data from the GCS we used, in addition to country criterion, individual characteristics such as size, sector, and whether the firm exports or not. For those countries for which the report has no information we used the country average calculated by KPMG in their "Corporate tax rate survey".

[12] Recognizing that the dependent variable is one of many individual data sources entering in the regression, we lag the corruption measure and use the 2002 version.

[13] Taken from the database of political institutions constructed by Beck et. al. (2001)

[14] See Governance Matters III: Governance Indicators for 1996-2002 for a more thorough discussion and for presentation of regressions results.

[15] A recent statement of this critique can be found in Glaeser, La Porta, Lopez-de-Silanes, and Shleifer (2004), who assert that much of the correlation between subjective measures of governance and levels of development is attributable to this type of bias.

[16] We do not consider higher values for the true correlation than 0.6. This is because we are trying to see the extent to which halo effects might result in an observed correlation of

0.8 which is substantially higher than the true correlation. If the observed correlation and the true correlation are close to each other, then the halo effects argument becomes unimportant empirically.

[17] In Kaufmann, Kraay and Zoido-Lobaton (1999), Table 5, we show that the estimated margins of error would be roughly twice as large if we assume that the correlation of error terms across sources is 0.5 instead of 0.

[18] For treatments of these effects in survey data, see Kaufmann and Wei (1999) and Hellman, Kaufmann and Schankerman (2000)

[19] For example, in our discussions with PRS, we learned that this source penalizes rich countries that in their view have the resources to reduce corruption but fail to do so.

[20] It is of course possible that halo effects are associated with countries' recent growth performance, rather than with income levels. We can use the analysis of this section to consider this case as well. The main insight is that since the correlation between recent growth and governance is typically fairly modest, growth-related halo effects would not need to be as large in order to impart a proportionately larger bias to this correlation.

[21] See for example Hall and Jones (1999), Acemoglu, Johnson and Robinson (2000), Kaufmann and Kraay (2002), Alcala and Ciccone (2004), Rodrik, Subramanian and Trebbi (2004), and Rigobon and Rodrik (2004).

[22] We use their specification excluding democracy, which implies that a one standard deviation increase in log per capita GDP improves rule of law by 0.14 standard deviations. They use a different measure of rule of law for the mid-1990s taken from Knack and Keefer (1995). However, its correlation with our rule of law indicator is above 0.8, so we can reasonably use the estimated coefficient from this paper with our

governance indicator, suitably standardized. Note also that in the system of equations estimated by Rigobon and Rodrik (2004) the conditional expectation of governance given per capita income also reflects the indirect effects of income on openness, which in turn affects the rule of law. However, these estimated indirect effects are so small that our conclusions are essentially unaffected by ignoring them.

References

Acemoglu, Daron, Simon Johnson, and James Robinson (2000), 'The Colonial Origins of Comparative Development', American Economic Review, 91 (5), 1369-1401.

Alcala, Francisco and Antonio Ciccone (2004), 'Trade and Productivity', Quarterly Journal of Economics, 119 (2), 613-646.

Beck, Thorsten, George Clarke, Alberto Groff, Phillip Keefer, and Patrick Walsh (2001), 'New Tools in Comparative Political Economy: The Database of Political Institutions', World Bank Economic Review, 15 (1), 165-176.

Clague, Christopher, Philip Keefer, Stephen Knack, and Mancur Olson (1999), 'Contract-Intensive Money: Contract Enforcement, Property Rights, and Economic Performance', Journal of Economic Growth, 4 (2), 185-211.

DiTella, Rafael and Ernesto Shargrodsky (2003), 'The Role of Wages and Auditing During a Crackdown on Corruption in the City of Buenos Aires', Journal of Law and Economics, 46 (1), 269-292.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2002), 'The Regulation of Entry', Quarterly Journal of Economics, 117 (1), 1-37.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer, (2003), 'Courts', Quarterly Journal of Economics, 118 (2), 453-517.

Glaeser, Edward, Rafael LaPorta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2004), 'Do Institutions Cause Growth?', Journal of Economic Growth, 9 (3), 271-303.

Golden, Miriam and Lucio Picci (2003), 'Proposal for a New Measure of Corruption, and Tests Using Italian Data', Manuscript, University of Bologna.

Hall, Robert E., and Charles Jones (1999), 'Why Do Some Countries Produce So Much More Output per Worker than Others?' Quarterly Journal of Economics, 114 (1), 83-116.

Hellman, Joel and Daniel Kaufmann (2003), 'The Inequality of Influence,' in J. Kornai and S. Rose-Ackerman (eds.), Building a Trustworthy State in Post-Socialist Transition, New York, NY: Palgrave MacMillan, pp. 100-118.

Hellman, Joel, Geraint Jones, Daniel Kaufmann and Mark Schankerman (2000), 'Measuring Governance, Corruption, and State Capture: How Firms and Bureaucrats Shape the Business Environment in Transition Economies', World Bank Policy Research Working Paper No. 2312, Washington, DC.

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobatón (1999a), 'Aggregating Governance Indicators', World Bank Policy Research Working Paper No. 2195, Washington, D.C. http://www.worldbank.org/wbi/governance/pubs/aggindicators.html.

Kaufmann, Daniel and Aart Kraay (2002), 'Growth Without Governance', Economia, 3 (1), 169-215.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi (2004), 'Governance Matters III: Governance Indicators for 1996, 1998, 2000, and 2002', World Bank Economic Review, 18 (2), 253-287.

Kaufmann, Daniel, and Shang-Jin Wei (1999), 'Does Grease Money Speed Up the Wheels of Commerce?', World Bank Policy Research Working Paper No. 2254, Washington, DC.

Knack, Steven and Philip Keefer (1995), 'Institutions and Economic Performance: Cross-Country Tests Using Alternative Measures', Economics and Politics, 7 (), 207-227.

Knack, Steve and Mark Kugler (2002), 'Constructing and Index of Objective Indicators of Good Governance', Manuscript, The World Bank.

Rigobon, Roberto and Dani Rodrik (2004), 'Rule of Law, Democracy, Openness, and Income: Estimating the Interrelationships', Manuscript, MIT and Kennedy School.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi (2004), 'Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development', Journal of Economic Growth, 9 (2), 131-165.

Sachs, Jeffrey, John W. McArthur, Guido Schmidt-Traub, Margaret Kruk, Chandrika Bahadur, Michael Faye, and Gordon McCord, (2004), 'Ending Africa's Poverty Trap', Brookings Papers on Economic Activity, 2004(1), 117-240.

Figure 2.1: The Development Dividend of Good Governance

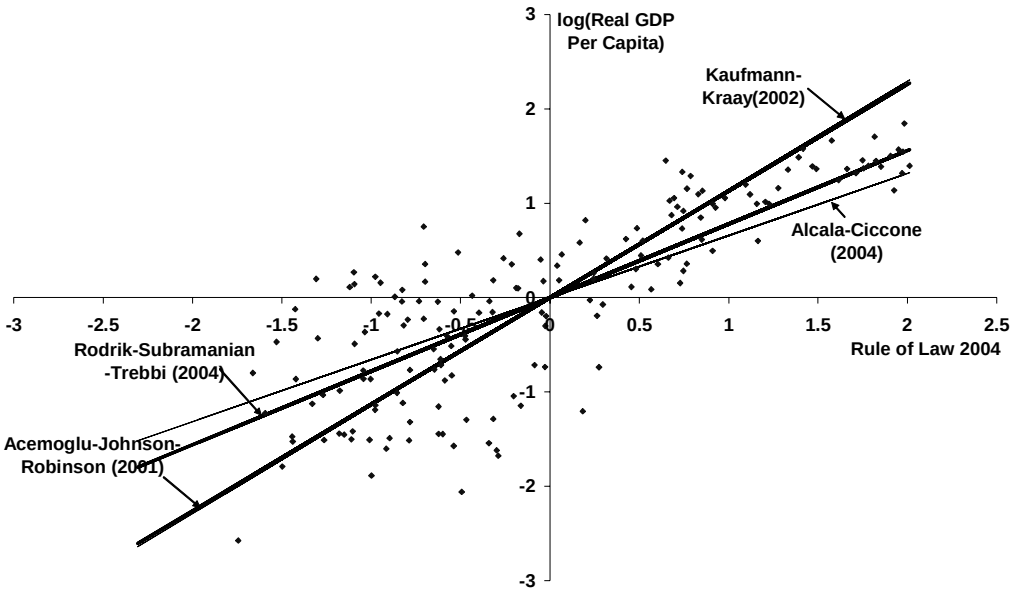
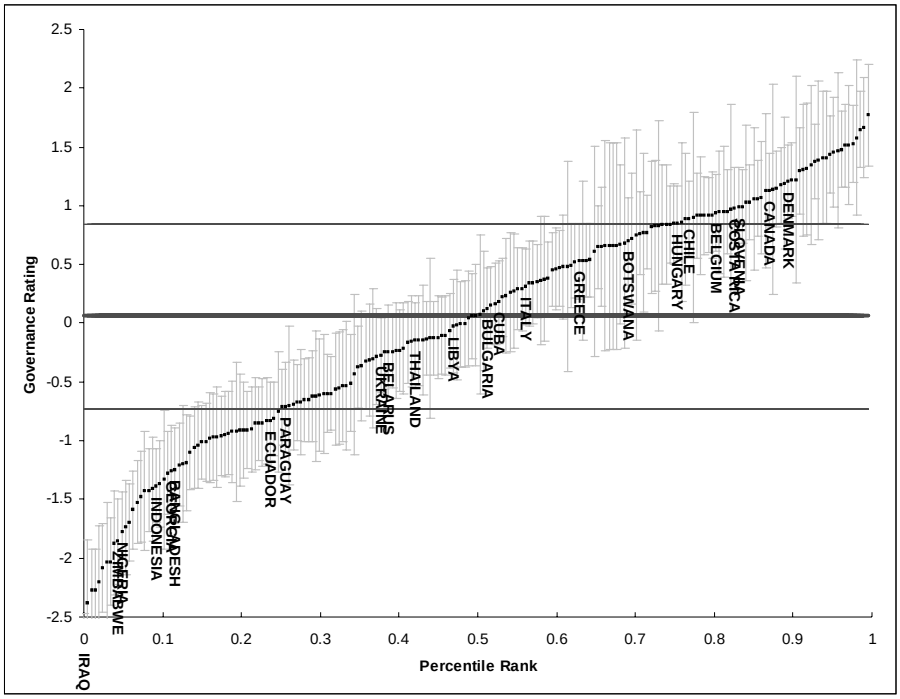


Figure 2.2: Margins of Error for Governance Indicators, 2004

Political Stability and Absence of Violence



Control of Corruption

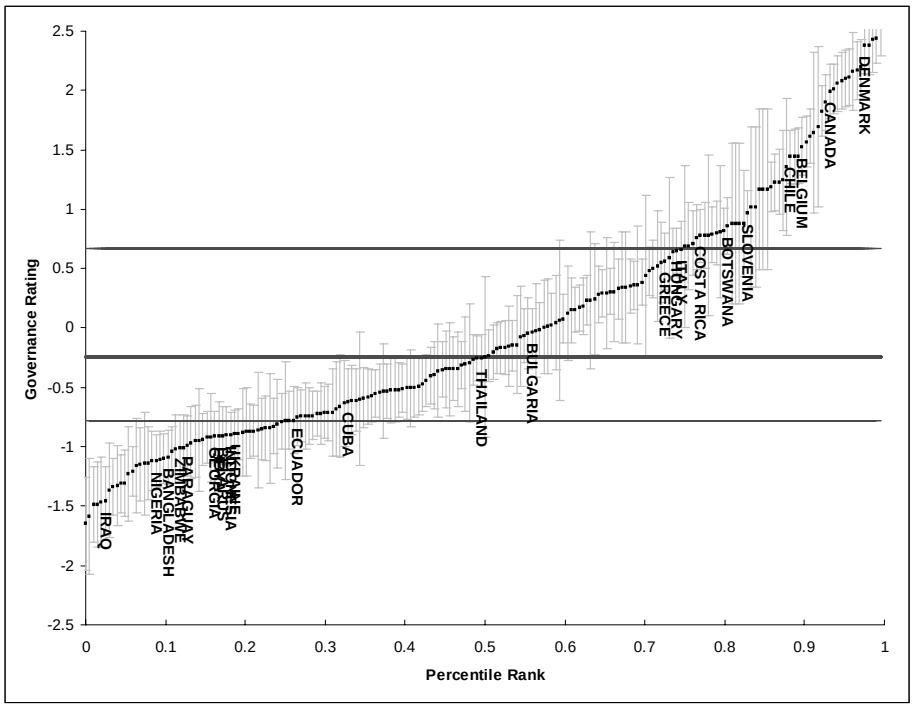
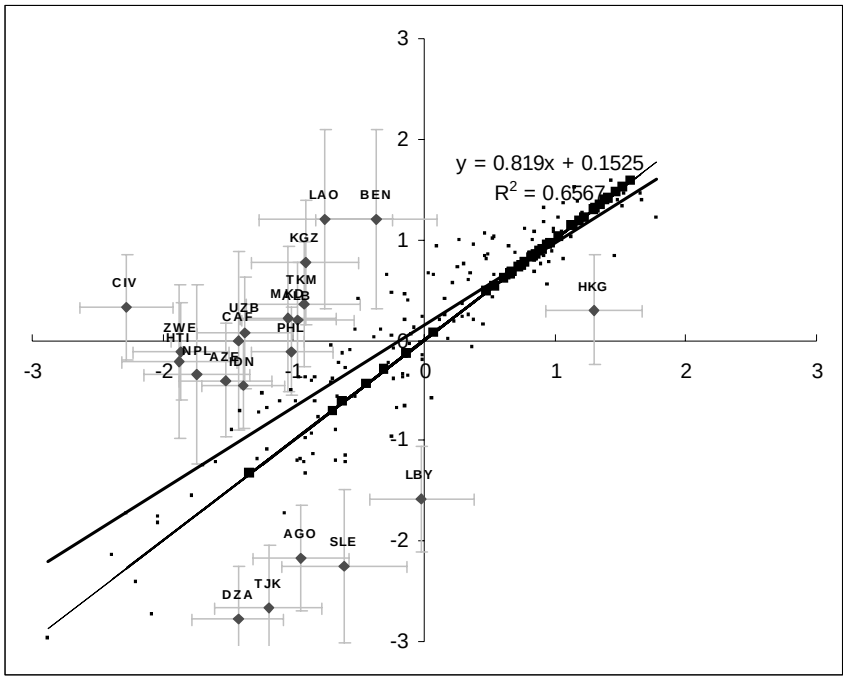


Figure 2.3: Changes Over Time in Governance Indicators 1996-2004

Political Stability and Absence of Violence (horizontal axis: 2004, vertical axis: 1996



Control of Corruption (horizontal axis: 2004,vertical axis: 1996

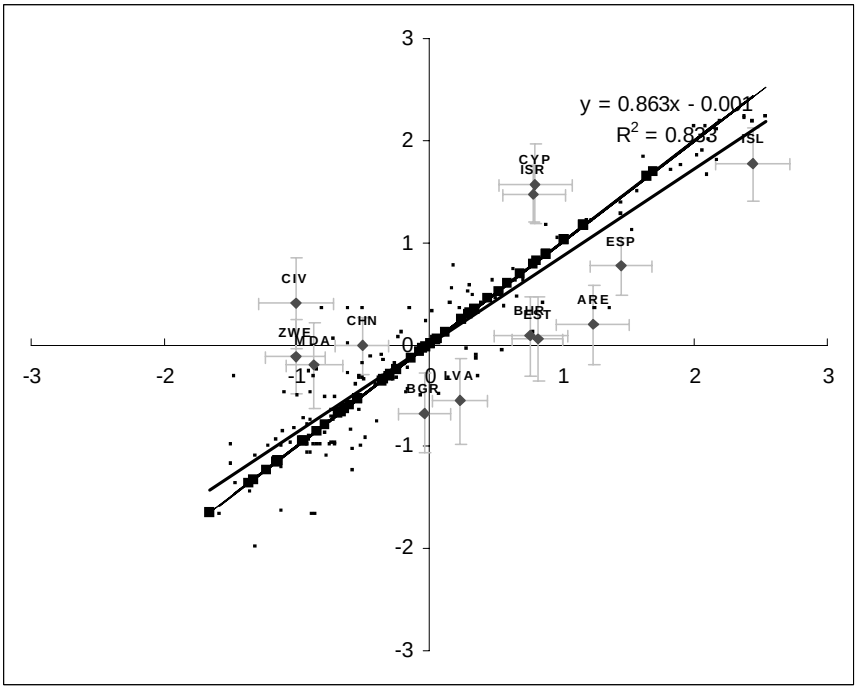


Figure 2.4: Significance of Changes in Individual Measures of Governance

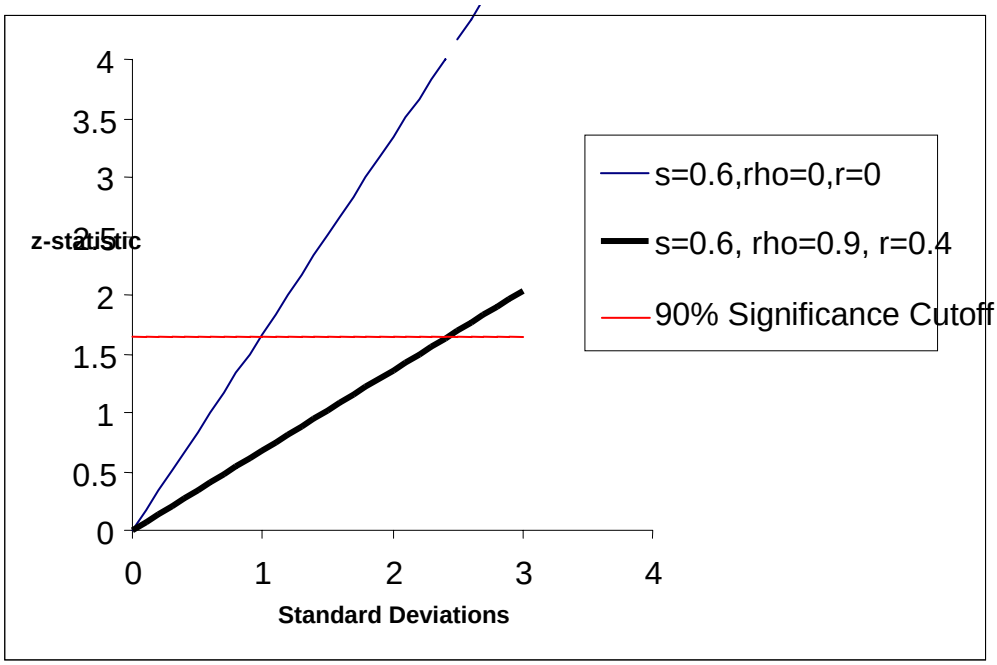


Figure 2.5: Changes in Measures of Ease of Business Entry, 2003-2004

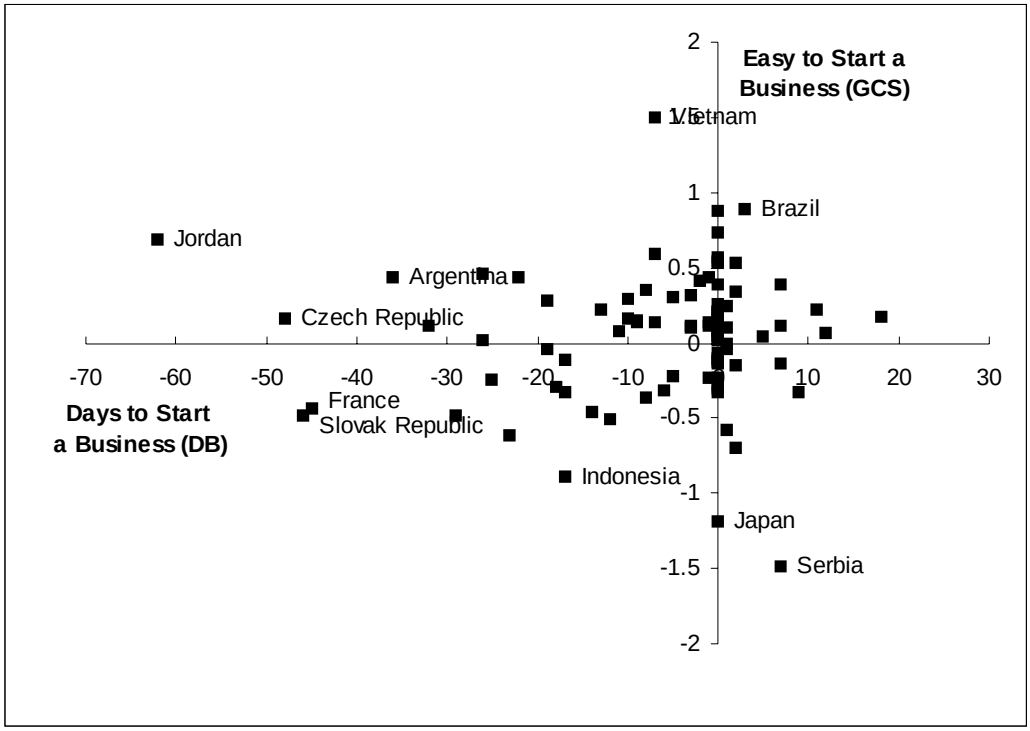


Figure 2.6: Halo Effects

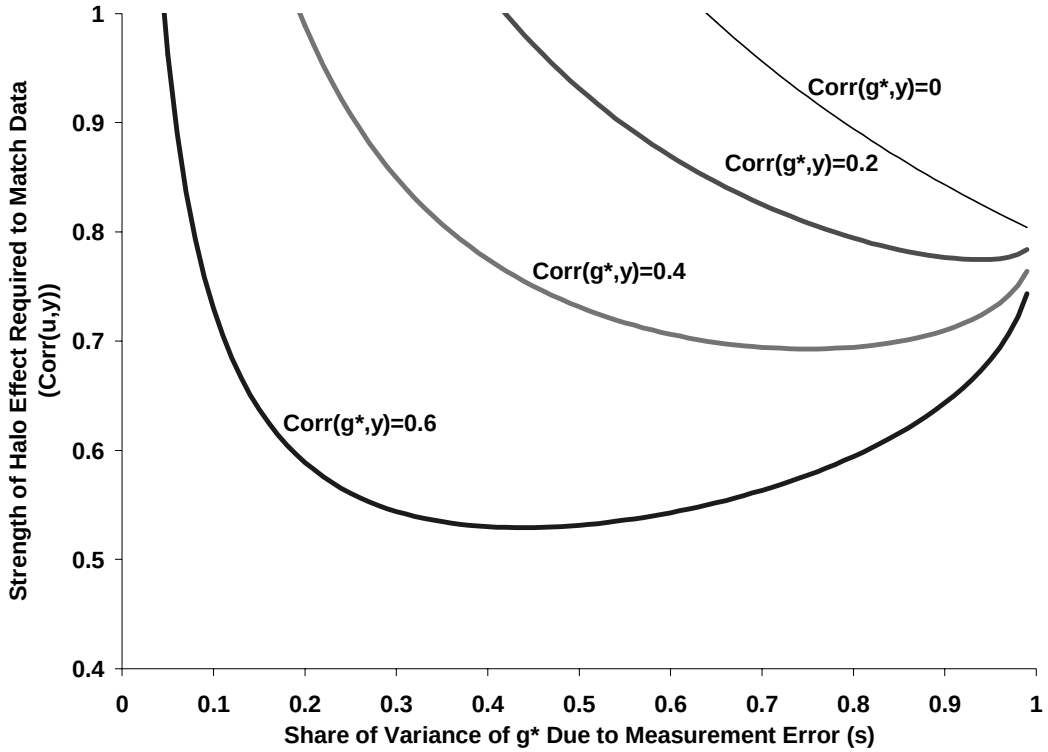


Figure 2.7: Governance and Per Capita Incomes in Africa

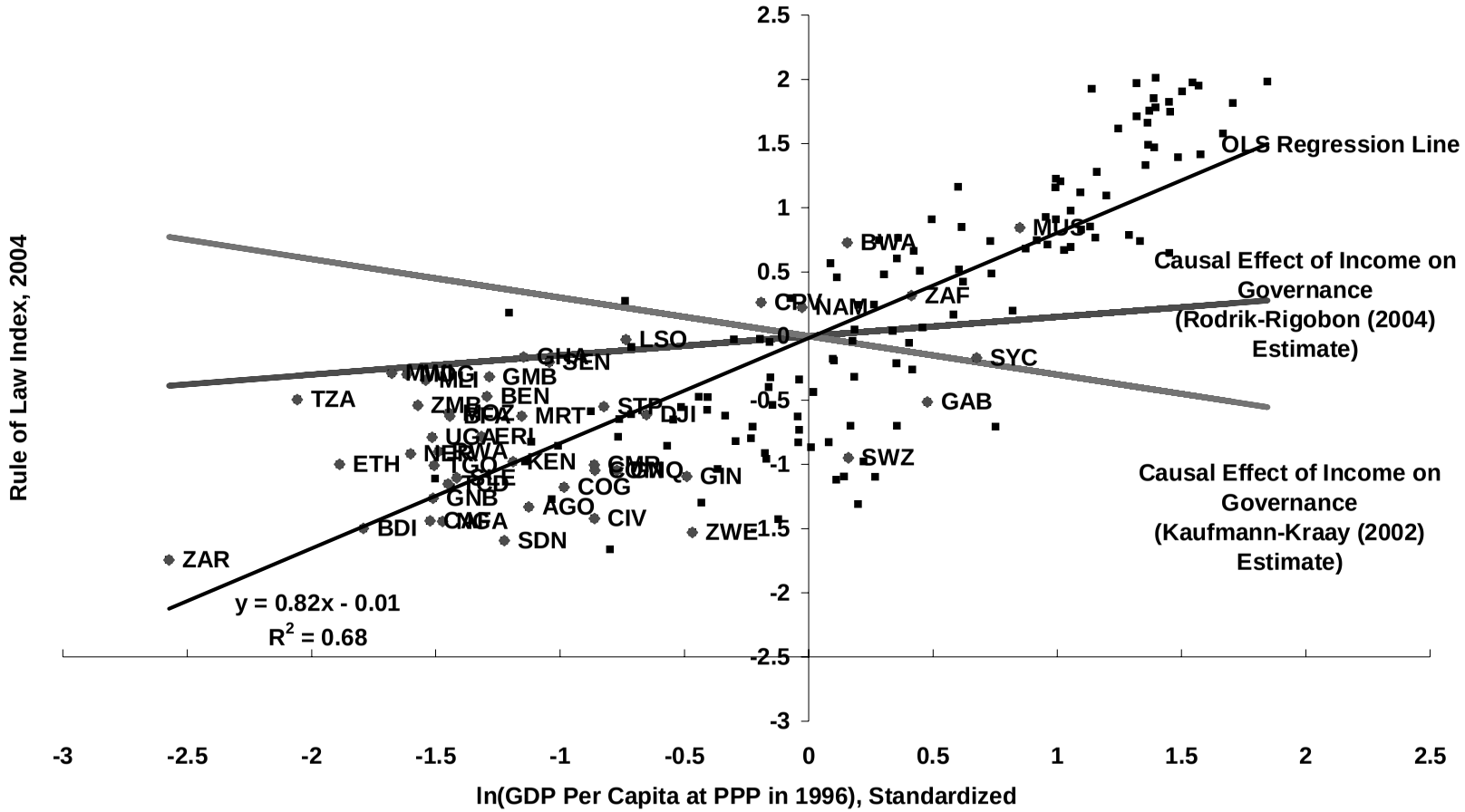


Table 2.1: Global Trends in Governance 1996-2004

[Quasi-Balanced Sample]* ** * †**

	# of Countries	World Average					Std. Dev. Across Countries					t-statistic for mean difference 1996-2004
		1996	1998	2000	2002	2004	1996	1998	2000	2002	2004	
Voice and Accountability												
EIU	115	0.41	0.42	0.42	0.46	0.46	*0.30*	*0.32*	*0.31*	*0.28*	*0.28*	1.5
PRS *	140	0.63	0.63	0.63	0.63	0.65	*0.25*	*0.27*	*0.27*	*0.26*	*0.26*	0.7
GCS **	88	..	..	..	0.49	0.51	..	..	..	*0.14*	*0.14*	….
FRH (PR+CL)	190	0.56	0.58	0.58	0.60	0.62	*0.34*	*0.33*	*0.33*	*0.33*	*0.33*	1.7
FRH (Press Freedom)	188	0.54	0.53	0.54	0.55	0.53	*0.24*	*0.26*	*0.25*	*0.25*	*0.25*	-0.2
WMO	186	..	..	..	0.55	0.53	..	..	..	*0.26*	*0.22*	…
Political Stability												
EIU	115	0.54	0.51	0.56	0.54	0.56	*0.29*	*0.30*	*0.30*	*0.28*	*0.26*	0.7
PRS *	140	0.78	0.71	0.73	0.75	0.75	*0.15*	*0.20*	*0.17*	*0.14*	*0.13*	-1.5
GCS **	88	0.73	0.74	0.63	0.66	0.67	*0.14*	*0.16*	*0.18*	*0.17*	*0.13*	-2.5 [-2.4]†
WMO	186	..	..	..	0.67	0.56	..	..	..	*0.24*	*0.20*	…
Government Effectiveness												
EIU	115	0.39	0.45	0.44	0.38	0.38	*0.30*	*0.24*	*0.24*	*0.29*	*0.30*	-0.2
PRS *	140	0.63	0.66	0.57	0.64	0.62	*0.19*	*0.12*	*0.11*	*0.17*	*0.15*	-0.4
GCS **	82	0.48	0.52	0.53	0.54	0.55	*0.22*	*0.28*	*0.27*	*0.24*	*0.23*	1.9 [2.8]†
WMO	186	..	..	..	0.56	0.55	..	..	..	*0.23*	*0.22*	…
Regulatory Quality												
EIU	115	0.42	..	..	0.51	0.55	*0.25*	..	..	*0.25*	*0.23*	4.3
GCS **	82	0.43	0.44	0.38	0.31	0.35	*0.15*	*0.15*	*0.16*	*0.14*	*0.13*	-3.4 [-3.0]†
WMO	186	..	..	..	0.58	0.61	..	..	..	*0.25*	*0.17*	…
HERITAGE ***	155	0.50	0.48	0.49	0.50	0.50	*0.30*	*0.31*	*0.31*	*0.29*	*0.28*	0.0
Rule of Law												
EIU	115	0.47	0.50	0.48	0.52	0.52	*0.27*	*0.30*	*0.30*	*0.26*	*0.26*	1.4
PRS *	140	0.72	0.62	0.65	0.62	0.63	*0.23*	*0.26*	*0.23*	*0.24*	*0.22*	-3.4
GCS **	82	0.67	0.63	0.57	0.51	0.51	*0.18*	*0.24*	*0.25*	*0.22*	*0.22*	-4.6 [-2.9]†
WMO	186	..	..	..	0.58	0.57	..	..	..	*0.23*	*0.20*	…
HERITAGE ***	155	0.50	0.48	0.46	0.44	0.44	*0.30*	*0.31*	*0.31*	*0.29*	*0.28*	-1.8
QLM	115	0.45	0.45	0.46	0.46	0.45	*0.29*	*0.29*	*0.30*	*0.30*	*0.30*	0.1
Control of Corruption												
EIU	115	0.35	0.34	0.33	0.35	0.35	*0.31*	*0.32*	*0.31*	*0.32*	*0.33*	0.2
PRS *	140	0.59	0.51	0.47	0.41	0.42	*0.21*	*0.21*	*0.21*	*0.19*	*0.19*	-7.2
GCS **	82	..	0.66	0.69	0.64	0.66	..	*0.29*	*0.25*	*0.22*	*0.21*	0.0 [-0.1]†
WMO	186	..	..	..	0.52	0.54	..	..	..	*0.27*	*0.20*	…
QLM	115	0.39	0.40	0.40	0.40	0.38	*0.29*	*0.29*	*0.29*	*0.29*	*0.29*	-0.2

Note that all variables are scaled to run from 0 to 1

* PRS Country coverage in 1996: 129 countries, all other periods 140.

** GCS Country coverage in 1996: 58; in 1998: 59; in 2000: 75; and in 2002 and in 2004: 82.

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

† Values in square brackets for GCS report t-stats for fully balanced sample from 1996 (same 52 countries)

Table 2.2: Classifying Countries for the MCA

	Control of Corruption			*WMO*	*DRI*	*GCS*
	2004	2000	1996	2004	2004	2004
Probability of Being Above the Median Is:						
			Number of Countries			
Below 10%	17	15	16	10	5	3
Below 25%	24	24	19	17	11	6
Between 25% and 75%	20	20	18	38	11	12
Above 75%	26	25	15	15	12	12
Above 90%	23	22	11	6	7	8
Total Number of Countries	70	69	52	70	34	30
			Proportion of Countries			
Below 10%	0.24	0.22	0.31	0.14	0.15	0.10
Below 25%	0.34	0.35	0.37	0.24	0.32	0.20
Between 25% and 75%	0.29	0.29	0.35	0.54	0.32	0.40
Above 75%	0.37	0.36	0.29	0.21	0.35	0.40
Above 90%	0.33	0.32	0.21	0.09	0.21	0.27
Average Standard Error	0.18	0.25	0.35	0.41	0.42	0.44

Table 2.3: Imputed Margins of Error for Objective Governance Indicators

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

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

Table 2.4: *De Jure* and *De Facto* Measures

Dependent Variable is GCS '04: "Easy to Start a Business?"	**1**	**2**	**3**	**4**	**5**
All Countries					
# of Days to start business (DB '04)	-1.18	-0.43	-0.47	-0.60	-0.59
	5.46***	1.87*	1.96*	4.33***	4.19***
Corporate Tax Rate			-0.01		0.01
			1.06		0.69
Control of Corruption (2002)		0.47	0.45	0.18	0.18
		6.14***	5.84***	2.80***	2.81***
Administrative Regulations (GCS '04)				0.75	0.77
				9.86***	9.05***
Observations (# of countries)	**81**	**81**	**81**	**81**	**81**
Adjusted R-squared	**0.23**	**0.44**	**0.44**	**0.71**	**0.71**
Developing Countries					
# of Days to start business (DB '04)	-0.49	-0.32	-0.29	-0.49	-0.47
	1.44	0.95	0.86	2.42**	2.25**
Corporate Tax Rate			0.01		0.01
			0.66		0.73
Control of Corruption (2002)		0.50	0.53	0.19	0.22
		3.30***	3.08***	1.48	1.67
Administrative Regulations (GCS '04)				0.83	0.82
				8.76***	8.73***
Observations (# of countries)	**56**	**56**	**56**	**56**	**56**
Adjusted R-squared	**0.01**	**0.19**	**0.18**	**0.57**	**0.57**
OECD + Newly-Industrialized Countries					
# of Days to start business (DB '04)	-0.97	-0.53	-0.57	-0.73	-0.74
	3.29***	1.65	1.88*	3.41***	3.33***
Corporate Tax Rate			-0.04		0.00
			1.92*		0.09
Control of Corruption (2002)		0.75	0.62	0.29	0.29
		2.85***	2.38**	1.28	1.25
Administrative Regulations (GCS '04)				0.64	0.65
				4.44***	3.51***
Observations (# of countries)	**25**	**25**	**25**	**25**	**25**
Adjusted R-squared	**0.18**	**0.36**	**0.46**	**0.69**	**0.67**

Note: DB refers to "Doing Business" study, GCS refers to Global Competitiveness Survey

Table 2.4, Cont'd: De Jure and De Facto Measures

Dependent Variable is GCS '04: "How Heavy Is Overall Tax Burden?"

	1	2	3	4	5
All Countries					
# of Days to start business (DB '04)			-0.96		-0.27
			0.46		0.15
Corporate Tax Rate	0.29	0.28	0.27	0.18	0.18
	2.37**	2.29**	2.22**	1.58	1.55
Control of Corruption (2002)		-0.77	-0.96	0.58	0.52
		1.27	1.19	0.91	0.62
Administrative Regulations (GCS '04)				-4.29	-4.28
				3.91***	3.91***
Observations (# of countries)	**81**	**81**	**81**	**81**	**81**
Adjusted R-squared	**0.09**	**0.09**	**0.08**	**0.24**	**0.23**
Developing Countries					
# of Days to start business (DB '04)			-2.06		-1.46
			0.68		0.54
Corporate Tax Rate	0.11	0.02	0.01	0.03	0.02
	0.71	0.15	0.09	0.16	0.11
Control of Corruption (2002)		-2.66	-2.80	-1.59	-1.71
		1.78*	1.88*	1.07	1.16
Administrative Regulations (GCS '04)				-2.93	-2.87
				1.62	1.60
Observations (# of countries)	**56**	**56**	**56**	**56**	**56**
Adjusted R-squared	**0.00**	**0.04**	**0.03**	**0.09**	**0.08**
OECD + Newly-Industrialized Countries					
# of Days to start business (DB '04)			0.96		2.37
			0.35		0.93
Corporate Tax Rate	0.63	0.64	0.64	0.33	0.32
	4.59***	4.43***	4.52***	3.56***	3.90***
Control of Corruption (2002)		0.47	0.78	2.63	3.49
		0.23	0.32	1.70	1.94*
Administrative Regulations (GCS '04)				-5.15	-5.38
				4.54***	4.89***
Observations (# of countries)	**25**	**25**	**25**	**25**	**25**
Adjusted R-squared	**0.47**	**0.45**	**0.42**	**0.65**	**0.65**

Note: DB refers to "Doing Business" study, GCS refers to Global Competitiveness Survey

Appendix 1: Statistical Significance of Changes in Aggregate Indicators

In this Appendix we extend the discussion in Section 3 of the paper to the problem of making inferences about changes over time in country governance based on our aggregate indicators. We develop a two-period version of the unobserved components model that we have used to construct the aggregate indicators in each period. We then use it to be more precise about the statistical significance of changes over time in our estimates of governance.

Let y(j,k,t) denote the governance assessment provided by individual data source k in period t for country j. We use a two-period version of the unobserved components model to express this observed data as a linear function of unobserved governance in country j at time t, g(j,t), and an error term capturing the various sources of measurement error that we have been discussing, ε(j,k,t):

(6) $$y(j,k,t) = \alpha(k,t) + \beta(k,t) \cdot \left(g(j,t) + \varepsilon(j,k,t)\right)$$

The intercept and slope parameters $\alpha(k,t)$ and $\beta(k,t)$ vary by data source and over time. As in our single-period model we assume that unobserved governance and the error terms are normally distributed with mean zero. We maintain the identifying assumption that unobserved governance and the all the error terms are mutually independent, i.e. $E[g(j,t) \cdot \varepsilon(j,k,s)] = 0$ for all sources k and periods t and s, and $E[\varepsilon(j,k,t) \cdot \varepsilon(j,m,s)] = 0$ for all sources k different from m and for all periods t and s. We also maintain as a choice of units that the variance of unobserved governance is one in each period, i.e. $E[g(j,t)^2] = 0$ for all t. Our only substantive new assumption relative to the basic one-period unobserved components model that we use to construct our governance indicators is that

unobserved governance is correlated over time, as are the error terms, i.e. $E[g(j,t)\cdot g(j,t-1)]=\rho$, and $E[\varepsilon(j,k,t)\cdot\varepsilon(j,k,t-1)]=r_k\cdot\sigma(k,t)\cdot\sigma(k,t-1)$, where ρ and r_k are the correlations over time of governance and the error term in source k, respectively.

Next let y(j,t) denote the Kx1 vector of observed data for each country; $\alpha(t)$, $\beta(t)$, $\sigma(t)^2$ and r denote the Kx1 vectors of the parameters in period t; and let B(t), $\Sigma(t)$ and R denote KxK matrices with the vectors $\beta(t)$, $\sigma(t)^2$ and r on their diagonals. Then using the properties of the multivariate normal distribution, the joint distribution of unobserved governance in the two periods in a country, conditional on the observed data for that country is normal with mean and variance:

$$(7)\qquad E\left[\begin{array}{c|c} g(j,t) & y(j,t) \\ g(j,t-1) & y(j,t-1)\end{array}\right]=\begin{pmatrix}\iota' & \rho\cdot\iota' \\ \rho\cdot\iota' & \iota'\end{pmatrix}\Omega^{-1}B^{-1}\begin{pmatrix}y(j,t)-\alpha(t) \\ y(j,t-1)-\alpha(t-1)\end{pmatrix}$$
$$V\left[\begin{array}{c|c} g(j,t) & y(j,t) \\ g(j,t-1) & y(j,t-1)\end{array}\right]=\begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix}-\begin{pmatrix}\iota' & \rho\cdot\iota' \\ \rho\cdot\iota & \iota'\end{pmatrix}\Omega^{-1}\begin{pmatrix}\iota' & \rho\cdot\iota' \\ \rho\cdot\iota & \iota'\end{pmatrix}$$

where B is a block-diagonal matrix with B(t) and B(t-1) on the diagonal, and ι is a Kx1 vector of ones. The covariance matrix Ω has the following block form:

$\Omega=\begin{pmatrix}\Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22}\end{pmatrix}$, with $\Omega_{11}=\iota\iota'+\Sigma(t)$, $\Omega_{12}=\Omega_{21}'=\rho\iota\iota'+R\Sigma(t)^{1/2}\Sigma(t-1)^{1/2}$, and $\Omega_{22}=\iota\iota'+\Sigma(t-1)$.[23]

The conditional mean and variance in Equation (7) are just the two-period generalizations of the estimates of governance and their precision based on the one-period unobserved components model (see Equations (2) and (3) in Kaufmann, Kraay, and Mastruzzi (2004)). In fact, if we set $\rho=r_k=0$ for all sources k, then we recover exactly the estimates of governance that we had before. The advantage of this two-period

formulation is that we now have specified the joint distribution of governance in the two periods for each country, conditional on the observed data in the two periods. Since we have modeled the joint distribution over the two periods of governance, we can base inferences about governance in the two periods, as well as changes in governance, on this joint distribution. We also note that the discussion of inference about changes over time in governance based on individual indicators in the previous section is just a special case of this more general formulation.[24]

We implement this two-period model using our actual dataset, over the period 1996-2004. We restrict attention to a balanced set of sources that are available in both periods for the two indicators. In order to implement this calculation, we need to have estimates of the parameters of the model in both periods (the α's, β's, and σ's), as well as estimates of the correlation over time of the errors in the individual sources (the r's) and the correlation of unobserved governance itself, ρ. We obtain these parameters in two steps. First, we estimate the one-period unobserved components model in 1996 and in 2004, to obtain estimates of the α's, β's, and σ's. We refer to this as the "static model" estimates. We also retrieve the estimates of governance and standard errors from the static model, to use as a basis for comparisons with the two-period model. Second, we calculate the correlation over time of these static estimates of governance as an estimate of ρ. In this second step we also insert the estimated parameters of the static model into Equation (6) and retrieve estimates of the errors in the sources in the two periods as residuals. The correlation over time in these estimated residuals serves as our estimate of the correlation in the errors. We then insert all the estimated parameters, together with the data, into Equation (7) to obtain our final estimates of governance in the two periods

conditional on the data, as well as the variance-covariance matrix of these estimates. We refer to these as the "dynamic model" estimates.

Appendix Table 2.A.1 summarizes the results of this calculation for the six governance indicators. In the top panel we present some summary statistics to aid in the comparison of governance estimates based on the single-period, or static model, and the two-period, or dynamic model. In the first two columns we report the correlation between the estimates of governance based on the static and dynamic models, in the two periods, 2004 and 1996. These correlations are virtually one for all six indicators in both periods, suggesting that our estimates of the levels governance do not change very much if we take into account persistence in governance and in the error terms. The third column reports the correlation of the change over time in the estimates of governance according to the two models. In light of the high correlations in levels between the two models, it is not very surprising that the correlation of changes is also very high, averaging 0.93 across the six indicators.

The next two columns of Appendix Table 2.A. report the average absolute change in the governance estimates for the static and dynamic models. These changes are roughly half as large in the dynamic model than in the static model, averaging 0.17 and 0.32 respectively. The reason the dynamic model gives much smaller estimates of the change in governance over time is because the estimated persistence in governance is quite strong relative to the estimated persistence in the error terms. Averaging across the six indicators, the persistence in unobserved governance is estimated to be 0.89. This is over twice as large as the persistence in the error terms, which averages 0.42 across all sources and indicators. Based on our intuitions from the simple example above, we

should expect to find substantially smaller estimates of the change in governance when we take this pattern of persistence into account, and this is in fact what happens.

The bottom panel of Appendix Table 2.A.1 summarizes the consequences of this persistence for inference about changes in governance. Formally our objective is to test the null hypothesis that the change in unobserved governance is zero conditional on the observed data. We begin by calculating the z-statistic associated with this hypothesis for each country, using the static and dynamic models. For the static model, we simply take the absolute change in our estimate of governance, and divide by the square root of the sum of the variances of the estimate of governance in the two periods. For the dynamic model, we calculate the variance of the change in governance as the sum of the estimated variances in the two periods, minus twice the estimated covariance between the two periods. The square root of this variance becomes the denominator of the z-statistic for the dynamic model. The average z-statistics are smaller in the dynamic model than in the static model, again consistent with the intuitions developed above. For the static model, the z-statistics average 0.82, as opposed to 0.59 for the dynamic model. This in turn implies fewer statistically significant changes in governance based on the dynamic model, as reported in the next two columns. The average number of significant changes at the 10 percent level falls by half from 21 to 10 once we take persistence into account.

Although a relatively small number of changes in the aggregate indicators signal statistically significant changes in unobserved governance, it is worth noting that the proportion of significant changes is much higher for the aggregate indicator than it is for individual indicators. Recall from the previous subsection that only the top one percent of changes in an individual indicator with typical persistence in unobserved governance

and the error term would be significant at the 90 percent level. This is not because individual indicators do not register large changes for individual countries – in fact frequently they do so. Rather, it is because the margins of error associated with changes in individual data sources are large. In contrast, for the aggregate indicators we find that between five and seven percent of all changes signal statistically significant changes in governance at the same significance level, reflecting the greater precision of the aggregate indicators. This illustrates the benefits of aggregation for assessing changes over time, as well as levels, of governance.

Finally, it is useful to compare the statistically significant changes in governance identified by the dynamic model with the "large" changes in governance we identified in Section 2.3 of this paper using a very simple rule of thumb. We begin by identifying all changes in governance based on the static model for which the 90 percent confidence intervals in the two periods do not overlap, as per the rule of thumb. Note that this is a more stringent condition for identifying significant changes in governance than the t-tests for the static model we have just discussed.[25] On average, there are nine significant changes in governance per indicator according to this rule of thumb applied to the simple static model, as compared with 10 in the dynamic model. There is a remarkable degree of overlap between the significant changes identified by the rule of thumb and the dynamic model. On average, eight of the nine changes identified by the rule of thumb are also significant in the dynamic model. Moreover, comparing the second and third-last columns of this panel, it is clear that the dynamic model turns up very few significant changes not identified by the rule of thumb. Although the simple rule of thumb and the more formal model turn up more or less the same set of significant changes in

governance, it is important to note that the magnitude of these changes is substantially smaller in the formal dynamic model.

In summary, we have developed a dynamic version of the single-period unobserved components model that we have used to construct our aggregate governance indicators. The advantage of specifying a dynamic version of the model is that it allows us to make formal statistical inferences about changes in unobserved governance based on our changes in the composite governance indicators. But this advantage comes at a cost. The two-period model is substantially more complicated to implement, particularly when the set of underlying data sources is not the same in both periods. Given that the number of data sources we use has expanded substantially over time, this is a significant limitation. Fortunately, however, we have seen that using a simple rule of thumb for identifying large changes over time in our static or single-period estimates of governance corresponds quite closely to formal inference regarding the significance of changes in governance. Because of this, we continue to use the single-period unobserved components model to construct the aggregate governance indicators in each period, and recommend using the simple rule of thumb that 90 percent confidence intervals do not overlap for identifying changes in governance that are likely to be statistically significant.

Appendix Table 2.A.1: Persistence and Inference about Changes in Governance Over Time

Summary Statistics

	Correlations			Mean Absolute Changes		Persistence	
	Levels, 2004	Levels, 1996	Changes, 1996-2004	Static	Dynamic	Governance	Average for Source Errors
VA	1.00	0.99	0.96	0.27	0.14	0.93	0.39
PV	0.99	0.99	0.98	0.44	0.30	0.78	0.39
GE	0.99	0.99	0.93	0.27	0.11	0.92	0.35
RQ	0.99	0.99	0.93	0.36	0.21	0.86	0.36
RL	0.99	0.99	0.88	0.23	0.12	0.94	0.53
CC	0.99	0.99	0.92	0.33	0.16	0.89	0.50
Average	0.99	0.99	0.93	0.32	0.17	0.89	0.42

Consequences of Persistence for Inference

	Mean t-Statistics		Number Significant at 90%		Rule of Thumb	
	Static	Dynamic	Static	Dynamic	Number Significant	Also Significant in Dynamic Model
VA	0.85	0.57	26	13	12	12
PV	0.91	0.78	21	18	14	14
GE	0.69	0.41	12	1	1	1
RQ	0.86	0.63	25	14	11	9
RL	0.73	0.55	16	7	7	5
CC	0.90	0.58	26	7	10	7
Average	0.82	0.59	21	10	9	8

[23] To obtain Equation (7), note that the (2K+2)x1 vector $(g(t), g(t-1), y(t), y(t-1))'$ is normally distributed with mean $(0, 0, \alpha(t), \alpha(t-1))'$ and variance-covariance matrix V with the following block form: $V_{11} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $V_{12} = \begin{pmatrix} \iota' & \rho \cdot \iota' \\ \rho \cdot \iota' & \iota' \end{pmatrix} B$, and $V_{22} = B\Omega B'$. Standard results for the partitioned multivariate normal distribution imply that the distribution of governance conditional on the observed data is normal with mean and variance given by Equation (7).

[24] To see this, set the number of sources K=1 and assume that $\alpha(t)=0$, $\beta(t)=1$, and $\sigma(t)=\sigma$ for this one source. Equation (5) then gives the conditional mean and variance of the level of governance in the two periods based on this single source. The expected change in governance conditional on the data is then just the difference between the conditional means in the two periods, and the conditional variance of the change is just the sum of the variances in the two periods less twice the covariance.

[25] Requiring 90 percent confidence intervals not to overlap is equivalent to requiring the absolute change in estimated governance to be larger than the sum of the standard errors in the two periods. This sum is always larger than the square root of the sum of the squares of these standard errors.