

Aggregating Governance Indicators

Daniel Kaufmann
Aart Kraay
Pablo Zoido-Lobaton

With the right method, aggregate indicators can provide useful estimates of basic governance concepts as well as measures of the imprecision of these aggregate estimates and their components.

The World Bank
Development Research Group
Macroeconomics and Growth
and
World Bank Institute
Governance, Regulation and Finance
October 1999



Summary findings

In recent years the growing interest of academics and policymakers in governance has been reflected in the proliferation of cross-country indices measuring various aspects of governance.

Kaufmann, Kraay, and Zoido-Lobaton explain how a simple variant of an unobserved components model can be used to combine the information from these different sources into aggregate governance indicators. The main advantage of this method is that it allows quantification of the precision of both individual sources of governance data and country-specific aggregate governance indicators.

Kaufmann, Kraay, and Zoido-Lobaton illustrate the methodology by constructing aggregate indicators of bureaucratic quality, rule of law, and graft for a sample of 160 countries. Although these aggregate governance indicators are more informative about the level of

governance than any single indicator, the standard errors associated with estimates of governance are still large relative to the units in which governance is measured. In light of these margins of error, it is misleading to offer very precise rankings of countries according to their level of governance: small differences in country rankings are unlikely to be statistically—let alone practically—significant. Nevertheless, these aggregate governance indicators are useful because they allow countries to be sorted into broad groupings according to levels of governance, and they can be used to study the causes and consequences of governance in a much larger sample countries than previously used (see for example the companion paper by Kaufmann, Kraay, and Zoido-Lobaton, “Governance Matters,” Policy Research Working Paper 2196).

This paper—a joint product of Macroeconomics and Growth, Development Research Group; and Governance, Regulation, and Finance, World Bank Institute—is part of a larger effort in the Bank to study the causes and consequences of governance for development. Copies of this paper are available free from the World Bank, 1818 H Street, N.W., Washington, DC 20433. Please contact Diane Bouvet, Room J3-273, telephone 202-473-5818, fax 202-334-8350, email dbouvet@worldbank.org. This paper is also available on the Web at http://www.worldbank.org/wbi/governance/working_papers.htm. The authors may be contacted at dkaufmann@worldbank.org, akraay@worldbank.org, or pzoidolobaton@worldbank.org. October 1999. (39 pages)

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent.

Aggregating Governance Indicators

Daniel Kaufmann
Aart Kraay
Pablo Zoido-Lobaton

The World Bank

October 1999

Abstract: In recent years, the growing interest of academics and policymakers in governance has been reflected in the proliferation cross-country indices measuring various aspects of governance. In this paper we explain how a simple variant of an unobserved components model can be used to combine the information from these different sources into aggregate governance indicators. The main advantage of this method is that it allows us to quantify the precision of the both individual sources of governance data as well as the aggregate governance indicators. We illustrate the methodology by constructing aggregate indicators of bureaucratic quality, rule of law, and graft, for a large sample of 160 countries. Although these aggregate governance indicators are more informative about the level of governance than any individual indicator, the standard errors associated with estimates of governance are still large relative to the units in which governance is measured.

The World Bank, 1818 H Street N.W., Washington, D.C. 20433.

(dkaufmann@worldbank.org, akraay@worldbank.org, pzoidolobaton@worldbank.org).

The views expressed in this paper are the authors' alone, and in no way reflect those of the World Bank, its Executive Directors, or the countries they represent. We would like to thank without implication Craig Burnside, Brad Efron, Eduardo Engel, Gil Mehrez, Jakob Svensson, Scott Wallsten, Shang-jin Wei, and seminar participants at the World Bank, UNDP and Latin American Econometric Society Meetings for helpful comments.

1. Introduction

In recent years, the growing interest of academics and policymakers in the extent, causes and consequences of governance and misgovernance has been reflected in the proliferation of cross-country indicators of various aspects of governance. In an accompanying paper (Kaufmann, Kraay, Zoido-Lobaton (1999)) we present a large database compiling several hundred cross-country indicators of various aspects of governance, produced by thirteen different organizations, and covering 178 countries. These indicators report subjective perceptions on a wide range of issues relating to governance, ranging from the extent to which corruption in the political system affects foreign investment, to the efficiency of public services delivery, to the likelihood that citizens of a country resort to extrajudicial means to settle disputes.

In this paper, we take the view that many of these indicators serve as imperfect proxies for one of a much smaller number of fundamental concepts of governance. Given this view, there are considerable benefits from combining related indicators into a small number of aggregate governance indicators. First, the aggregate indicators span a much larger set of countries than any individual source, permitting comparisons of governance across a broad set of countries. Second, aggregate indicators can provide more precise measures of governance than individual indicators. Third, it is possible to construct quantitative measures of the precision of both the aggregate governance indicators and their components, allowing formal testing of hypotheses regarding cross-country differences in governance.

We realize these benefits by constructing aggregate governance indicators using an unobserved components model. This model expresses the observed data as a linear function of unobserved governance plus a disturbance term capturing perception errors and/or sampling variation in each indicator. The main advantage of this method is that it allows us to obtain estimates of the variance of this disturbance term for each indicator. These can be interpreted as a measure of how informative each indicator is about the broader concept of governance it measures. We then compute the mean of the conditional distribution of governance given the observed data for each country as a natural point estimate of the level of governance in that country. Similarly, the variance

of this conditional distribution provides a natural estimate of the precision of this aggregate governance measure for each country.

We illustrate our approach with reference to three fundamental aspects of governance: rule of law, government effectiveness, and graft. We group 31 indicators constructed in 1997 and/or 1998 into three clusters corresponding to these three concepts of governance, and compute aggregate indicators spanning 166, 156 and 155 countries respectively. In our companion paper documenting the governance database, we construct similar indices for several other aspects of governance.

Although the unobserved components methodology we use is quite standard, we find its application to the construction of composite governance indicators interesting.¹ One of our major findings is that the aggregate governance indicators we construct are rather imprecise, despite the high correlations observed between various sources of governance data. In particular, a 90% confidence interval around the point estimate of governance for a typical country spans almost the entire interquartile range of the distribution of estimated governance. This implies that although it is possible to robustly identify twenty or so countries with the best and worst governance in the world, it is much more difficult to identify statistically significant differences in governance among the majority of countries.

Our results are based on three key assumptions: (1) that the measurement errors in individual indicators of governance are uncorrelated across indicators; (2) that the relationship between unobserved governance and observed indicators is linear, and (3) that the distribution of unobserved governance across countries is normal. Relaxing the first assumption is difficult to do in practice, simply because without this assumption we cannot determine whether the correlation of observed scores across indicators is merely due to correlated perception errors or whether it reflects the common concept of governance being measured. However, under the likely alternative that perception errors are correlated across sources, the measures of precision we report will be biased downwards. As a result, the standard errors we report should be interpreted as a lower bound on the precision of aggregate governance indicators. We consider the

¹ Unobserved components models were pioneered in economics by Goldberger (1972), and the closely-related hierarchical and empirical Bayes models in statistics by Efron and Morris (1971, 1972).

consequences of relaxing the second assumption by proposing a method which simply aggregates the ordinal information across indicators. Although this has the advantage of simplicity and does not require assumptions of linearity, it is also much less precise than the unobserved components method since it discards the cardinal information in the data. The third assumption of a normal distribution for unobserved governance implies that our estimates of governance will be clustered around the mean of this distribution. This raises the possibility that the difficulty in distinguishing between countries is in part driven by this assumption. We therefore explore the robustness of our results by considering alternatives to this assumption, and find that our conclusions are materially unaffected by our assumptions on the shape of the distribution of unobserved governance.

These findings have several implications for policy and empirical research on the causes and consequences of governance for economic development. At a basic level, the finding that governance is imprecisely measured should warn against taking too seriously the exact point estimates of governance, as well as country rankings based on these estimates. At best, it is possible to sort countries into broad categories according to their levels of governance, and even then there is considerable uncertainty regarding the category to which many countries should be assigned. To emphasize this point, we avoid discussions of specific countries in this paper. Second, since available indicators of governance are noisy measures of “true” governance, empirical work which uses these indicators as explanatory variables may well underestimate the impact of governance due to the usual attenuation bias caused by badly-measured right-hand side variables. Since our methodology allows us to quantify the measurement errors in these variables, it is possible to obtain rough measures of the extent of this attenuation bias.² Finally, our results suggest that if we want to more precisely differentiate among countries according to their level of governance, we need to improve the quality and quantity of data gathered on governance.

The remainder of this paper proceeds as follows. In Section 2, we motivate the empirical work which follows by describing the indicators of governance we use to

² In our companion paper we explore this idea in more detail, using cross-country regressions of per capita income on various governance measures, instrumenting for governance using measures of the linguistic composition of the population -- in the spirit of Hall and Jones (1999).

illustrate our ideas. In Section 3, we lay out and implement the unobserved components framework for estimating governance, and present the main results in Section 4. In Section 5, we discuss the consequences of relaxing several of the assumptions underlying the model. We conclude with a discussion of the implications of our findings for research and policy advice regarding governance.

2. Indicators of Governance

In this paper, we use data from 31 different indicators of governance constructed in 1997 and/or 1998. These indicators are drawn from 13 different sources and are grouped into three clusters corresponding to rule of law, government effectiveness, and graft. The key features of these indicators are summarized in Table 1, and a detailed description of the sources and variables can be found in Kaufmann, Kraay and Zoido-Lobaton (1999). In the first two columns of Table 1 we identify each source of governance data by abbreviation and by name. In the next three columns, we report the source of information for each measure (surveys of residents or polls of experts), the country coverage, and a measure of the extent to which the sample of countries covered by each indicator is representative of the population of countries in the world. In the remaining columns we report the specific concepts measured by each source in each of the three clusters.³

A quick look at Table 1 shows that these indicators differ along several dimensions. First, even within clusters there is considerable variation in the particular concept measured by each indicator. For example, questions about graft range from the incidence of “improper practices” (WCY) to the likelihood that additional payments are required to “get things done” (WDR). Similarly, questions regarding the rule of law range from whether citizens can successfully sue the state to whether citizens are likely to resolve disputes extra-judicially. Despite this heterogeneity, we take the view that within each cluster, each of these concepts is an imperfect indicator of the corresponding broader concept of governance.

The second respect in which these indicators differ is in the nature of the respondents who provide the information. Slightly less than half of the indicators are surveys of businesspeople and/or residents of a country, while the remaining indicators are polls of experts who rate a set of countries according to various criteria. As we discuss in more detail later in the paper, this difference between these two types of

³ For a number of these sources, we use the average of several questions relating to the corresponding core concepts of governance. As we discuss subsequently, we are reluctant to include individual questions from a single source separately in our analysis, as the necessary assumption that measurement errors are uncorrelated across indicators is much more difficult to support for the case of multiple questions from a single source.

indicators has implications for how we interpret the error terms in the relationship between observed indicator scores and the underlying concepts of governance.

The third respect in which these indicators differ is in the sample of countries they cover. A number of indicators cover a very large and broad sample of developed and developing countries (EIU, DRI, HFWSJ, PRS and WDR), while others cover very narrowly-focused samples of countries (PERC for Asia, CEER and FHNT for transition economies). Some indicators cover primarily developed countries but also include major developing countries (WCR, GALLUP, BERI). This difference between indicators is perhaps the most important for the empirical work which follows. There is by now considerable evidence that governance on average tends to be better in richer countries. This implies that the distribution of governance is likely to be very different in indicators which cover sets of countries with different average income levels. These differences need to be taken into account when placing the observed data from various indicators into common units and combining them into aggregate governance indicators.

In order to distinguish between indicators in this dimension, we construct a simple coverage index which measures differences between the distribution of countries across income and regional classifications and the distribution of all countries in the world across these categories. In particular, we divide the world into a two-way classification by region and income, following the World Bank's 1998 World Development Report. For each of the sources of governance data, we report one-half of the sum of absolute deviations between the share of countries in each of the 45 region/income categories in that source and in the world as a whole. By construction, this measure ranges from zero to one, with low values indicating more representative indicators. We report this number in the fifth column of Table 1. The five indicators covering the largest number of countries (DRI, EIU, HF, PRS and WDR) are substantially more representative according to this measure than the others, with a value of the coverage index of less than 0.25. In our subsequent empirical work we will refer to these as representative indicators, and the remainder as non-representative indicators.

Finally, we note that each of these sources of governance data uses different units to measure governance. Most polls of experts report discrete categorical responses (e.g. the prevalence of corruption on an integer scale from one to four), while

for most surveys of citizens or entrepreneurs we have the mean response across respondents of discrete categorical scores. We re-orient data from each source so that higher values correspond to better outcomes (i.e. stronger rule of law, more effective government, and less graft). In addition, we rescale each indicator by subtracting the minimum possible score and dividing by the difference between the maximum and minimum scores, so that each indicator is on a possible scale from zero to one. Since we rescale each indicator using the maximum and minimum possible scores (rather than the maximum and minimum actual scores in the sample of countries covered by each indicator), this is nothing more than a convenient choice of units.

In Table 2, we report the pairwise correlations among indicators within each of the three governance clusters. The great majority of these are positive and substantial, frequently greater than 0.6. In the empirical work which follows, we will interpret these large correlations within clusters as reflecting the common component of governance in these indicators. It is interesting to note that despite the strong pairwise correlations among these indicators, and despite the favourable interpretation that these correlations reflect the common component of governance rather than correlated perception errors, we nevertheless find that governance is not very precisely measured. We provide some intuitions for this in the following section.

3. Estimating Governance

In this section we interpret the data as being generated by an unobserved components or multiple-indicator model in which the observed data on governance can be expressed as a linear function of unobserved governance plus a random error term. We review the well-known features of this model, and propose a simple extension which delivers consistent parameter estimates for representative as well as non-representative indicators. We then describe how the parameters of this model can be estimated and can be used to construct estimates of each of the three aspects of governance in each country.

The Model

Our data consists of clusters of indicators of three aspects of governance – rule of law, government effectiveness, and graft. Let $g(j)$ denote an unobserved index of one of these three aspects of governance in country j , for example, graft. The observed data on graft consists of a cluster of $k=1, \dots, K$ indicators, each one providing a numerical rating of some aspect of graft in each of the $j=1, \dots, J(k)$ countries covered by that indicator. We assume that we can write the observed score of country j on indicator k , $y(j,k)$, as a linear function of unobserved governance, $g(j)$, and a disturbance term, $\varepsilon(j,k)$, as follows:

$$(1) \quad y(j,k) = \alpha(k) + \beta(k) \cdot (g(j) + \varepsilon(j,k))$$

where $\alpha(k)$ and $\beta(k)$ are unknown parameters which map unobserved governance $g(j)$ into the observed data $y(j,k)$. We assume that $g(j)$ is a random variable with mean zero and variance one. Our objective is to summarize our knowledge about $g(j)$ for each country j using the distribution of $g(j)$ conditional on the observed data $y(j,k)$, $k=1, \dots, K(j)$ for country j . The mean of this conditional distribution provides a natural estimate of the level of governance in country j , and the variance of this conditional distribution is a natural measure of the precision of this indicator of governance. The assumption of a zero mean and unit variance for governance is an innocuous choice of units required to identify the parameters $\alpha(k)$ and $\beta(k)$. Since we will allow the variance of the error term

to vary across indicators k , the fact that $\beta(k)$ multiplies the error term is an innocuous rescaling which slightly simplifies some of the expressions which follow.

We use this unobserved components model, which treats unobserved governance as a random variable rather than as a fixed parameter to be estimated, for a pragmatic reason. We will shortly also assume that the variance of the disturbance term $\varepsilon(j,k)$ may differ across indicators. In this case, we cannot treat the $g(j)$ s as fixed parameters to be estimated for each country, since individual effects are not identified in a fixed effects model with heteroskedastic disturbances.⁴ Moreover, it should be clear from Equation (1) that naive aggregates such as a simple average of rescaled indicators for each country will not result in sensible estimates of governance, as long as the parameters $\alpha(k)$ and $\beta(k)$ differ across indicators and different countries appear in different sets of indicators. It is also not possible to remove the dependence of the observed data on these nuisance parameters by standardizing (i.e. by removing the sample mean from each indicator, and dividing by the sample standard deviation). This is because if indicator k is non-representative, the sample mean will reflect not only $\alpha(k)$, but also the mean of $g(j)$ in the sample of countries covered by indicator k . Even if an indicator is representative in the sense that the standard deviation of unobserved governance is equal to one in the sample of countries it covers, the sample standard deviation of observed scores will reflect not only $\beta(k)$, but also the standard deviation of the disturbances.

The disturbance term $\varepsilon(j,k)$ captures two sources of uncertainty in the relationship between true governance and the observed indicators. First, the particular aspect of governance covered by indicator k is imperfectly measured in each country, reflecting either perception errors on the part of experts (in the case of polls of experts), or sampling variation (in the case of surveys of citizens or entrepreneurs). Second, the relationship between the particular concept measured by indicator k and the corresponding broader aspect of governance may be imperfect. For example, even if the particular aspect of graft covered by some indicator k , (such as the prevalence of

⁴ To see this, consider the special case where $\alpha(k)=0$ and $\beta(k)=1$ for all indicators. We can make the likelihood function of the observed data arbitrarily large simply by estimating $g(j)$ as the observed score on a particular indicator, for example $g(j)=y(j,K)$, for every country j , and setting $\sigma_\varepsilon(K)=0$. Kiefer (1980) provides a detailed discussion of this point.

“improper practices”) is perfectly measured, it may nevertheless be a noisy indicator of graft if there are differences across countries in what “improper practices” are considered to be.

We assume that the disturbance term has zero mean, $E[\varepsilon(j,k)]=0$; has the same variance across countries within a given indicator but a different variance across indicators, $E[\varepsilon(j,k)^2]=\sigma_{\varepsilon}(k)^2$; and is independent across indicators and countries, $E[\varepsilon(j,k)\varepsilon(j',k')]=0$ if $j \neq j'$ or $k \neq k'$. The variance of the error term can be interpreted as a measure of how informative indicator k is about $g(j)$, and is likely to vary across indicators. The assumption that the errors are independent across indicators is a strong one, but unfortunately one that is also difficult to relax. Intuitively, without this assumption we cannot identify whether the correlation of scores between two indicators is due to their common component of governance $g(j)$, or whether it simply reflects the correlation of errors. In contrast, this identifying assumption maintains that all of the correlation of scores across indicators is attributable to their common estimate of governance. We will consider the consequences of relaxing this assumption in the following section. For now, we simply note that this identifying assumption corresponds to a “best case” scenario regarding the precision of governance aggregates, since it assumes that each indicator provides independent information on a particular aspect of governance. As a result, we are if anything likely to overstate the precision with which governance is measured.

The parameters $\alpha(k)$ and $\beta(k)$ map unobserved governance into the observed data. Although all of our indicators (after rescaling) are nominally in the same units and are measured on a scale from zero to one, there are nevertheless three reasons why these parameters may differ across indicators. First, not all indicators use the entire range of possible scores. For example, although WDR measures perceptions of graft on a scale from one to six, the lowest observed score is only 2.36. This suggests that $\alpha(k)$ on this indicator may be greater than that of an indicator such as EIU which uses the full range of possible scores. Second, a given indicator might be “easy” (“tough”) relative to other indicators in the sense that it tends to overestimate (underestimate) a particular aspect of governance in countries where it is in fact low (high). This would be reflected in a relatively high (low) value of $\alpha(k)$ on that indicator. Third, consider a non-representative indicator that covers a set of countries in which the average level of a

particular aspect of governance is better than in the world as a whole (e.g. BERI, which covers primarily developed countries). Suppose further that this source tends to score countries relative to each other, so that the worst (best) country in the sample receives the lowest (highest) possible score of zero (one). This would be reflected in a relatively high value of $\beta(k)$, since relatively small differences in true governance are magnified into relatively large differences in observed scores.

The Conditional Distribution of Governance

Our objective is to summarize our knowledge about governance in each country j by the distribution of governance conditional on the observed data in country j . This task is greatly simplified by assuming that both $g(j)$ and the disturbances $\varepsilon(j,k)$ are jointly normally distributed. In this case, $g(j)$ and $y(j,k)$, $k=1, \dots, K(j)$ are jointly normal, and the conditional distribution of $g(j)$ given the data is also normal, with mean and variance given by:

$$(2) \quad E[g(j)|y(j)] = \iota'(\iota'\iota + \Sigma)^{-1} B^{-1} \cdot (y(j) - \alpha)$$

$$(3) \quad V[g(j)|y(j)] = \iota - \iota'(\iota'\iota + \Sigma)^{-1} \iota$$

where $y(j)$ is a $K(j) \times 1$ vector which stacks the $K(j)$ data points for country j , α is the corresponding $K(j) \times 1$ vector of $\alpha(k)$ s, B and Σ are $K(j) \times K(j)$ diagonal matrices with the corresponding $\beta(k)$ s and $\sigma_{\varepsilon}(k)^2$ s on the diagonal, and ι is a $K(j) \times 1$ vector of ones. We refer to the conditional mean in (2) as the estimated value of that aspect of governance in country j . With a slight abuse of terminology, we refer to an interval from the $(\delta/2)^{\text{th}}$ percentile to the $(1-\delta/2)^{\text{th}}$ percentile of the conditional distribution of $g(j)$ as a δ -percent “confidence interval” around this estimate, and we refer to the square root of the conditional variance in (2) as the “standard error” of this estimate.⁵

⁵ This framework has a distinctly Bayesian interpretation. The distribution of $g(j)$ conditional on the observed data $y(j)$ can be viewed as a posterior distribution, and the mean of this distribution as an estimator of $g(j)$ would be justified as a point estimate of $g(j)$ by a quadratic expected posterior loss function. Similarly, the “confidence interval” is analogous to a Bayesian highest posterior density interval.

These expressions have a very natural interpretation. If the parameters $\alpha(k)$, $\beta(k)$ and $\sigma_\varepsilon(k)^2$ were known, a sensible way to estimate $g(j)$ would be to rescale the observed scores by subtracting $\alpha(k)$ and dividing by $\beta(k)$, and then construct a weighted average of these re-scaled scores. In particular, let $\tilde{y}(j,k) = \frac{y(j,k) - \alpha(k)}{\beta(k)} = g(j) + \varepsilon(j,k)$ denote the rescaled value of $y(j,k)$. Then the conditional mean in (2) is a weighted average of these standardized scores for country j on each of the $K(j)$ indicators in which it appears, with weights corresponding to the inverse of the variance of the error term on each indicator, i.e.

$$E[g(j)|y(j)] = \frac{\sum_{k=1}^{K(j)} \sigma_\varepsilon(k)^{-2} \cdot \tilde{y}(j,k)}{1 + \sum_{k=1}^{K(j)} \sigma_\varepsilon(k)^{-2}}. \quad \text{The conditional variance is simply}$$

$$V[g(j)|y(j)] = \left(1 + \sum_{k=1}^{K(j)} \sigma_\varepsilon(k)^{-2} \right)^{-1}, \quad \text{which is decreasing in the number of indicators}$$

available for that country, $K(j)$, and is increasing in the variance of the error term in each of these indicators, $\sigma_\varepsilon(k)^2$.

Estimating the Unknown Parameters

In order to implement (2) and (3), we need to first estimate the unknown parameters $\alpha(k)$, $\beta(k)$ and $\sigma_\varepsilon(k)^2$ for every indicator k . For the set of representative indicators, we can use the assumption of normality of $g(j)$ and $\varepsilon(j,k)$ to write down the likelihood function of the observed data. Provided that we have at least three such indicators, the model is identified and it is straightforward to maximize this function with respect to the $\alpha(k)$ s, $\beta(k)$ s, and $\sigma_\varepsilon(k)^2$ s to obtain estimates of the unknown parameters for the representative indicators.⁶

⁶ Although maximum likelihood estimation of these parameters requires the assumption of normality, it is also possible to dispense with this assumption and apply a method of moments procedure. In the just-identified case of three indicators, these methods lead to identical parameter estimates. In the overidentified case of more than three indicators, these methods differ only in the weights applied to the various moment conditions, and in practice this makes little difference for the parameter estimates.

We cannot apply this method to non-representative indicators. To see why, consider the maximum-likelihood estimate of $\alpha(k)$, which unsurprisingly is the mean score across countries on indicator k . It is straightforward to see from Equation (1) that the expected value of the sample mean of scores on indicator k is $\alpha(k) + \beta(k) \cdot \bar{g}(k)$, where $\bar{g}(k)$ denotes the average level of governance in the sample of countries covered by indicator k . For representative indicators, our choice of units for governance normalizes $\bar{g}(k) = 0$. However, for a non-representative indicator where the average level of governance is different from the world as a whole, $\bar{g}(k) \neq 0$ and the sample mean does not provide a consistent estimate of $\alpha(k)$.

We can nevertheless obtain consistent estimates of the unknown parameters by using the following simple argument. If $g(j)$ were observable, we could estimate $\alpha(k)$, $\beta(k)$ and $\sigma_\varepsilon(k)$ for any indicator by regressing the observed scores $y(j,k)$ on $g(j)$. Although $g(j)$ is itself not observable, we do have an estimate of $g(j)$ based on the representative indicators. In particular, let $g^*(j)$ denote the mean of $g(j)$ conditioning only on the data from the representative indicators. We can decompose this conditional mean into observed governance $g(j)$ plus its deviation from the mean $u(j)$, i.e. $g^*(j) = g(j) + u(j)$. Since $u(j)$ is independent of $g(j)$, we can view $g^*(j)$ as measuring $g(j)$ with error, i.e. as a classic errors-in-variables problem. It is well-known that OLS estimates of $\beta(k)$ from a regression of $y(j,k)$ on $g^*(j)$ will produce downward-biased estimates due to the usual attenuation bias imparted by measurement error in $g^*(j)$. In particular, the probability limit of the OLS slope coefficient is $\beta(k) \cdot \left(1 - \frac{V[u(j)]}{V[g^*(j)]}\right)$. Since the variance of $u(j)$ is simply the variance of the conditional mean of $g(j)$ given in Equation (3), and since $V[g^*(j)]$ is observable, we can correct the OLS coefficients for this attenuation bias to arrive at consistent estimates of the parameters of the non-representative indicators.⁷

⁷ An alternative approach to the problem of non-representative indicators would be to impute data for the missing observations (in the spirit of Rubin (1987)). We do not pursue this approach here simply because it is difficult to specify the key ingredient of the imputation process – the conditional distribution of the unobserved data given the observed data – in our application.

It is worth noting that this estimation method “rewards conformity”, in the sense that indicators that are highly correlated will have low estimated variances and hence will be perceived as more precise. Given our assumption that the disturbance terms are independent across indicators, it makes sense to treat highly correlated indicators in this way. If on the other hand indicators are correlated simply because their disturbances are correlated, this interpretation would be inappropriate. We take this issue up in more detail in Section 4, and argue that it will result in even less precise estimates of governance than those we obtain here.

4. Results

In this section we implement the unobserved components model laid out in the previous section for three concepts of governance: government effectiveness, rule of law, and graft. We first present estimates of governance and associated standard errors for each country, and then consider the consequences of these standard errors for identifying cross-country differences in governance. We conclude with a simple example which relates the pairwise correlations observed among individual governance indicators directly to the measures of precision of the aggregate indicators.

Estimates of Governance

Our main finding is that the available data do not permit very precise estimates of governance. We illustrate this point in Figure 1. In each of the three panels of Figure 1, we order countries on the horizontal axis by their estimate of governance, and on the vertical axis we plot the corresponding point estimate of governance, i.e. the conditional expectation of $g(j)$ given the observed data for country j , and a 90-percent confidence interval around this point estimate, i.e. the 5th and 95th percentiles of the conditional distribution of governance for each country j . The size of these confidence intervals varies by country, reflecting the fact that different countries appear in different numbers of sources, and that different countries appear in different sets sources of differing precision. To provide a sense of the dispersion in the point estimates of governance, the three horizontal lines in each graph delineate the quartiles of the distribution of the point estimates of governance for each cluster.

The most striking feature of Figure 1 is that these confidence intervals are large relative to the units in which governance is measured. For example, for a typical country the standard deviation of the conditional mean of rule of law or graft is around 0.3, so that a typical 90% confidence interval extends 0.5 above and below the point estimate of graft. In the case of government effectiveness, the standard deviation of the conditional mean is on average slightly larger and equal to 0.33, so that a 90% confidence interval extends 0.55 above and below the point estimate of rule of law. These confidence intervals are large in the sense that they are comparable in size to the entire interquartile range of the distribution of estimates of governance. Moreover, it should be noted that

these confidence intervals do not reflect the sampling variation in the point estimates of the unknown parameters $\alpha(k)$, $\beta(k)$ and $\sigma_\varepsilon(k)$. If this uncertainty were also taken into account, the standard errors would be even larger.

The parameter estimates reported in Table 3 reveal some interesting differences across indicators. To interpret the estimates of the $\alpha(k)$ s and $\beta(k)$ s, note that our assumption of a standard normal distribution for governance implies that the vast majority of countries will have governance ranging from -2 to 2 . Since the observed data range from zero to one, one might expect that a representative indicator would have $\alpha(k)=0.5$ and $\beta(k)=0.25$. Interestingly, there are significant departures from this benchmark. Several indicators (e.g. WDR) have estimated values of $\beta(k)$ substantially lower than this benchmark, and higher values of $\alpha(k)$, indicating that they do not use the entire range of possible scores. There is also a great deal of variation in the estimates of the standard deviation of the errors on each individual indicator, $\sigma_\varepsilon(k)$, suggesting that the precision with which individual sources measure governance varies widely.

Assessing Cross-Country Differences in Governance

An advantage of this methodology is that it permits straightforward tests of hypotheses regarding cross-country differences in governance. However, the large size of the confidence intervals documented in Figure 1 suggests that it will be difficult to find statistically significant differences in governance between many pairs of countries. We illustrate this point with two simple exercises. Suppose first that for each of the three aspects of governance, we want to group countries into quartiles according to their level of governance. A natural way to do this is to group countries according to their point estimates of governance, i.e. according to the mean of the conditional distribution of governance in each country. Moreover, a natural way to assess the confidence with which countries are assigned to quartiles is to consider the corresponding 90% confidence intervals shown in Figure 1. In particular, if the 90% confidence interval for a country falls entirely within a given quartile, the probability that this country in fact belongs in another quartile is less than 10%. For a small group of countries at each end of the distribution of governance, we can conclude with a great deal of confidence that these countries are in fact in the top and bottom quartiles. However, for the middle

quartiles the situation is much less clear, as very few countries' 90% confidence intervals lie entirely within a given quartile, for each of the three aspects of governance.

Clearly, the number of countries we can assign to a particular quartile using this rule depends on the size of the confidence interval. If we instead consider shorter confidence intervals, such as 75% or 50% intervals, we can better discriminate among countries, albeit with lower confidence. We explore this possibility in Table 4, where we report the proportion of all countries for which an $x\%$ confidence interval falls entirely within the indicated quartile, for the three governance aggregates in turn. We consider three possibilities, $x=90\%$, $x=75\%$ and $x=50\%$. At all significance levels, a substantial fraction of countries in the top and bottom quartiles can be clearly identified as belonging in these groups. As the size of the confidence interval declines, more and more countries can be significantly assigned to quartiles. Nevertheless, even at very low significance levels, only one-quarter to one-half of the countries in the middle two quartiles have confidence intervals lying entirely within their respective quartiles.

A related issue is the significance of pairwise differences in governance. In particular, for every pair of countries in which our point estimate of governance in country j is greater than in country j' , we can investigate the hypothesis that country j in fact has better governance by computing the probability that $g(j) > g(j')$.⁸ For countries with similar point estimates of governance, this probability will be close to 0.5, while for countries far apart in the distribution of governance, this probability will approach one. To illustrate this point systematically, for each country j in the sample, we compute the probability that, conditional on the observed data for countries j and j' , $g(j) > g(j')$ for every comparator country j' . We then compute the proportion of comparator countries for which this probability is between 5% and 95%. This is analogous to counting the number of comparator countries for which a conventional test at the 10% significance level of the null hypothesis that governance is the same in these two countries cannot be rejected. We summarize the results of these pairwise comparisons in Figure 2. We again order countries in ascending order according to their point estimates of governance on the horizontal axis, and we plot this proportion of comparator countries

⁸ Since the $g(j)|y(j)$ and $g(j')|y(j')$ are jointly normal and independent by assumption, this calculation involves a straightforward integration of the area under a bivariate normal probability density function.

as dark points on the vertical axis. We also repeat the exercise, but instead report the larger proportion of countries for which this probability is between 25% and 75%, which corresponds to a test at the 50% significance level. This proportion is shown as a light dot in Figure 2.

Not surprisingly, at the two ends of the distribution there are significant differences between the level of governance in these countries and most other countries, especially at the 50% significance level. However, there is also a strong inverted U-shaped pattern in this graph, reflecting the fact that a large fraction of countries are clustered near the middle of the distribution of estimated governance, and it is relatively difficult to distinguish among such countries. In particular, for the “typical” country around the middle of the distribution of governance, governance is not significantly different from nearly half of all other countries in the world, at conventional significance levels.

Intuitions

Our finding that governance is imprecisely measured is somewhat surprising. After all, in Section 2 we documented that fact that the pairwise correlations among various governance indicators are substantial, and the identifying assumption of independent errors across indicators implies that the only source of this observed high correlation among indicators is the unobserved common component of governance. One might therefore easily conclude that governance is quite well measured and that it is straightforward to distinguish among countries’ governance using this data. We now illustrate why this intuition is misleading, unless the correlations in the observed data are very high indeed.

As a specific example, suppose that there are only three representative indicators associated with a particular governance concept, i.e. $K=3$, and that the pairwise correlations among the observed scores are all equal to ρ . It is straightforward to show that in this case, the estimated variance of the residual will be $\sigma_{\varepsilon}(k)^2 = \frac{\rho - 1}{\rho}$

for each of the surveys $k=1,2,3$.⁹ Inserting this into Equation (3), the variance of the aggregate governance indicator based on this hypothetical data will be the same for each country and is equal to $V[g(j)|y(j)] = \frac{1-\rho}{1+2\cdot\rho}$. To give an idea of the magnitude of the corresponding 90% confidence intervals, we superimpose them on the hypothetical distribution of governance in the upper panel of Figure 3, for various values of ρ . As ρ increases, the confidence intervals become shorter. However, for the correlations of around 0.75 typically observed in our governance data, this confidence interval remains large relative to the units in which governance is measured.

In the lower panel of Figure 3, we relate this to the significance of cross-country comparisons. A simple summary statistic is the proportion of countries whose true level of governance lies within the 90% confidence interval of a particular country. This proportion will depend on the location of the reference country, and on the correlations in the observed data. We plot this proportion for the median country and the country at the first quartile of the distribution of governance, for various values of ρ . For the observed correlation of indicators of around 0.75, the 90% confidence interval around the point estimate of governance for the median country encompasses the true level of governance in about half of all other countries in the world, and somewhat less for a country at the first quartile. Only if the observed correlations are very large is it possible to distinguish the median country from most other countries with a high degree of confidence.

⁹ To see this, it is only necessary to solve the system of nine equations relating the three sample means and the six unique elements of the sample covariance matrix of the indicators to their population counterparts and solve for the unknown parameters.

4. Extensions

In this section we consider how our results depend on three assumptions underlying the unobserved components model of the previous sections: that the disturbances are independent across indicators, that the mapping from unobserved governance into observed data is linear, and that the distribution of unobserved governance is normal. We find that the first two assumptions if anything overstate the precision with which governance is measured. Relaxing the third assumption does not materially affect our results.

Correlated Disturbances

In the previous section we assumed that the disturbances $\varepsilon(j,k)$ were independent across indicators. Intuitively, this assumption allowed us to attribute all of the observed correlation of scores across indicators to the common component of governance $g(j)$, and hence permitted us to identify the portion of the variation in scores across countries within each indicator due to measurement error. A consequence of this assumption is that any indicator which is not very correlated with the others was interpreted as having a large residual variance.

Although useful, the assumption that the errors are independent across indicators may not be valid, for at least three reasons. First, in the case of polls of experts, it is possible that the perceptions of experts who rank countries on a particular indicator are influenced by their knowledge of countries' rankings on other indicators. Second, the errors in surveys of residents might be correlated across countries if residents of a particular country have a tendency to systematically overstate regulatory and governance obstacles, due to a broad-based predisposition to report a worse situation than is objectively warranted.¹⁰ Finally, it is possible that perceptions of governance from various indicators are unduly influenced by a single event, such as a high-profile scandal which is not representative of the level of graft in that country.

¹⁰ See Kaufman and Zoido-Lobaton (1999).

Although it is not possible to statistically identify the correlation of the disturbances across sources, it is straightforward to see the consequences of positively-correlated errors for our results. If the errors are correlated across indicators, each additional indicator contributes less information to our estimate of governance. This will be reflected in the variance of the conditional distribution of governance in each country. In particular, it is straightforward to show that holding constant the variance of the residuals on each indicator, the variance of the conditional distribution of governance is increasing in the correlation between the errors on any two indicators.¹¹

We illustrate the practical consequences of this observation in Table 5. For each of the three aggregate indicators, we re-estimate the variance of the conditional distribution of governance, imposing a range of assumptions on the correlation of the disturbances. For the purposes of this example, we restrict ourselves to a set of three representative indicators (EIU, DRI and PRS), and also to the set of about 100 countries which appear in all three indicators.¹² As the assumed correlation among the errors rises from 0 to 0.5, the aggregate governance indicators become less precise, although the magnitude of the effect depends on the indicator (since the estimated variances of the disturbances change as well). In the case of government effectiveness, the standard error of the aggregate increases only slightly, from 0.32 to 0.35. In contrast, for rule of law the standard error doubles from 0.33 to 0.66.

It is difficult to adequately address the problem of correlated disturbances simply because it is not possible to separately identify the correlations between the errors. Nevertheless, it is useful to realize that the estimated standard errors associated with point estimates of governance are likely to be substantially understated under the assumption of independent errors. This reinforces our argument of the previous section that cross-country comparisons of the level of governance should be made with caution.

¹¹ If the positive correlations between the residuals differ across pairs of indicators, the relative magnitudes of the estimated variances will also be affected. In particular, suppose that two indicators of bureaucratic quality are highly correlated with each other, but not very correlated with the third. In the previous section we assumed that the errors were independent, and so the high correlation between the first two indicators implied that the variance of the errors was small on these indicators relative to the third. However, if we knew that the high correlation between the first two indicators was due to correlated errors, then the estimated variances on these indicators would be large relative to that of the third indicator. As a result, the relative rankings of countries might also be affected.

¹² We do this only for simplicity, since it allows us to report just one standard error per aggregate, which is the same for all countries.

Non-Linearities

In the previous section we assumed that the relationship between unobserved governance and observed indicator scores was linear. This assumption places strong restrictions on the units in which governance is measured in the various indicators in our sample. For example, consider an indicator such as Gallup which asks respondents how many cases of corruption there are among public officials, and offers the choice of four broad categories: “none”, “a few”, “many” and “a lot”. Our observed data consists of numerical scores on a scale of one to four corresponding to these categories. The assumption of a linear mapping from unobserved graft into observed scores implies that the difference in graft between a country with a score of 4 and one with a score of 3 (i.e. the difference between “a lot” and “many” cases of corruption) is the same as the difference between two countries with scores of 3 and 2 (i.e. the difference between “many” and “a few”). Given the somewhat vague response categories, it is not at all clear that the assumption of linearity is warranted. Moreover, even if these categories were equally-spaced according to some appropriate metric, the fact that the observed data are discrete while our unobserved governance indicator is continuous violates the assumption of a linear relationship between the two.¹³

Finally, the mere fact that indicators are non-representative may also contribute to a non-linear relationship between governance and observed scores. A number of the indicators in our sample cover primarily developed countries together with a few developing countries. It is possible that the developing countries in these indicators suffer from a “curse of inclusion” in the sense that they receive worse scores than they might otherwise have received simply because they are implicitly being compared with countries in which various aspects of governance are likely to be much better.¹⁴ For example, representative surveys such as DRI or PRS assign moderate scores of 55/100

¹³ The straightforward solution to this problem would be to rely on an ordered multinomial choice model with individual effects in the latent variable. However, for such a model to be identified, it would again be necessary to assume that the variances of the errors are identical across indicators.

¹⁴ This problem may be particularly acute for polls of experts who consider a large set of countries at once, since in contrast to surveys of residents, experts are much more likely to be aware of relative comparisons of countries.

and 3/6 to Mexico for graft, while on a much less representative survey such as BERI which covers primarily developed countries, Mexico receives a rather poor rating of 1/7.

What can be done about these non-linearities? A very general solution to this problem might be to combine only the ordinal information in each indicator, i.e. the relative rankings of countries within each indicator. In particular, one can think of a given indicator as providing a ranking of pairs of countries according to their level of governance. The information in the relative rankings of countries from various indicators can then be combined by noting that if several indicators consistently rank country A as having better governance than country B, this provides evidence that governance is in fact better in country A than in country B. Clearly, this approach has several advantages. First, it is computationally very simple. Second, we do not have to know the choice of units in which governance is measured, or whether indicators are representative or not. Third, it does not require any assumption of linearity in the relationship between governance and observed scores. However, this method will result in larger standard errors for pairwise comparisons since it discards information in differences in the level of scores across countries and indicators.¹⁵

To illustrate the relative imprecision of such an ordinal aggregate, for every pair of countries j and j' on every indicator k , we construct an indicator variable $x(j,j',k)$ which takes the value 1 if country j is ranked higher than country j' with respect to the aspect of governance covered by indicator k , and zero otherwise.¹⁶ A natural null hypothesis to test is that governance in countries j and j' is the same, i.e. that the probability $x(j,j',k)$ is equal to one is 0.5. Under the assumption that the errors are independent across

¹⁵ A further drawback of this method is that it is difficult to construct an aggregate ranking of countries according to governance. One possibility would be to average the indicator variables $x(j,j',k)$ over all surveys k and partner countries j' , and rank countries according to this index. However, it is difficult to put standard errors on such a ranking, since even if the errors are independent across surveys, the $x(j,j',k)$ will not be independent across partner countries j' . A deeper problem with this method is that there is a fundamental result from social choice theory which places strong restrictions on the properties such an aggregate ranking may have. According to Arrow's Impossibility Theorem, it is impossible for any aggregation of each indicators' "preferences" to simultaneously satisfy three intuitive and desirable properties: (i) the aggregation respects unanimity -- if every indicator says that A is more corrupt than B, then so should our aggregate, and (ii) the aggregation displays the independence of irrelevant alternatives property -- the ranking of A and B does not depend on any indicators' ranking of A or B relative to any other country C; and (iii) the aggregation is non-dictatorial -- the aggregate ranking of A and B is not uniquely determined by a single indicator's ranking of A and B. In particular, Arrow's theorem tells us that if (i) and (ii) hold, then (iii) does not hold.

¹⁶ For those surveys which report discrete categorical scores, we discard "ties" as uninformative about the relative level of governance.

indicators, this hypothesis can be tested using the data on the proportion of indicators in which country j is ranked higher than country j' in a simple binomial proportions test.

We report the results of this exercise in Figure 4, which is analogous to Figure 2. We again order countries in ascending order according to their point estimates of governance on the horizontal axis, and we plot the proportion of all comparator countries for which the null hypothesis that governance in these two countries is equal cannot be rejected at the 10% significance level as dark dots on the vertical axis. The light dots report the same information, but at the 50% level. Comparing Figures 2 and 4, it is clear that the ordinal aggregate allows use to identify far fewer statistically significant differences in governance across countries. In fact, for many countries it is impossible to reject the null at the 10% level that governance in this country is the same as for every other country in the world using this method! Although this ordinal method is a useful vehicle for making rough comparisons across countries and requires little in the way of assumptions on the underlying data, it is much more difficult to obtain statistically significant differences among countries.

Alternative Distributions for Governance

In Section 2 we assumed that unobserved governance and the disturbances were jointly normally distributed. As we noted, this assumption has a significant payoff in terms of analytical tractability, as it ensured that the distribution of governance conditional on the observed data was normal, with simple expressions for its mean and variance. However, given the bell-shape of the normal distribution, this approach embodies the implicit assumption that a relatively large fraction of countries in the world have similar moderate levels of governance, and relatively few have either very good or very bad governance. There are two reasons to question this assumption. First, it is not at all clear a priori that this provides an accurate depiction of the true cross-country distribution of governance. Second, it is possible that our finding that it is difficult to statistically distinguish differences in governance between a large proportion of countries in the world is accentuated by the assumption of normality, which forces a disproportionate fraction of countries to be clustered near the mean of the distribution of governance. If instead we assumed that governance was more dispersed, then it is

possible that it is easier to identify statistically significant differences in governance across countries.

It is not clear how one might identify the shape of the true distribution of governance across countries, since it is difficult to disentangle the shape of this distribution from the shape of the distribution of the error terms. However, it is possible to explore the robustness of the results to different choices for the distribution of governance itself. We do this by instead assuming that unobserved governance follows a Beta[a,b] distribution. We consider three choices of parameters corresponding to three very different shapes of the possible distribution of governance. These three possibilities are illustrated in the left-hand column of Figure 5. We first consider $a=b=5$, which generates a symmetric bell-shaped distribution. This case serves as a benchmark in that it is very similar to the normal distribution we have been using so far. We also consider the possibility that the distribution of governance is skewed to the right ($a=2, b=5$), with relatively few countries with very good governance in the right tail of the distribution. Finally, we consider the possibility that governance is uniformly distributed ($a=b=1$), with a similar proportion of countries at each possible level of governance. We continue to assume that the distribution of the disturbances is normal.

On the right-hand side of Figure 5, we explore the consequences of these alternative assumptions for our conclusions about the significance of cross-country differences in governance. As a specific example, we focus on an aggregate of the three largest representative indicators of graft (EIU, DRI and PRS), and again restrict ourselves to the sample of about 100 countries appearing in all three indicators. For each country, we report the point estimates and standard errors corresponding to each assumption on the distribution of governance.¹⁷ As the assumed shape of the distribution of true governance changes, not surprisingly so does the distribution of point estimates. The more important observation is that our results on the difficulty of distinguishing between countries do not change. It is clear from Figure 5 that the

¹⁷ We compute these as follows. First, using a method of moments argument we construct estimates of the parameters of the model corresponding to the assumed distribution of governance (note that the mean and variance of this beta distribution change as we vary the parameters). We then construct the joint distribution of $g(j)$ and $y(j)$ as the appropriate mixture of a normal and a beta distribution, and then obtain the marginal distribution of $y(j)$ and conditional distribution of $g(j)$ given $y(j)$ by appropriate numerical integrations of this joint distribution. We numerically evaluate the mean of this distribution, and the 5th and 95th percentiles, and report these for each country in Figure 5.

number of countries with 90% confidence intervals falling entirely within particular quartiles is essentially unchanged for each quartile, as we vary our assumptions on the shape of the distribution of governance.

5. Conclusions

In this paper, we have taken the view that the many different available indicators of governance provide imperfect signals about a relatively small number of fundamental aspects of governance, such as rule of law, government effectiveness, and graft. We grouped the many available indicators into three clusters corresponding to these concepts of governance, and used a linear unobserved components model to obtain aggregate estimates of these three aspects of governance. Despite several optimistic assumptions, we find that governance is not very precisely measured using these aggregate indicators. In particular, although it is possible to identify statistically significant differences between countries at opposite ends of the distribution of governance, it is much more difficult to discriminate among the majority of countries with any degree of confidence.

Nevertheless, we find the aggregate governance indicators to be useful for several reasons. First, they are based on a methodology which provides a consistent framework for placing data from various sources into common units, taking into account that the samples of countries included in different sources may not be representative of the world as a whole. Second, the aggregate indicators span a much larger sample of 155 or more countries, permitting (admittedly-impure) comparisons across a much larger set of countries than is possible using any single indicator. Third, although the aggregate indicators are not as precise as one might have hoped, they are nevertheless much more reliable than any individual indicator. Finally, we believe that it is useful to have quantitative measures of the precision of aggregate indicators in order to caution users of both individual and aggregate indicators of the substantial margins of error associated with cross-country comparisons of governance.

Empirical research on governance issues can also benefit from the aggregate indicators presented here. Many empirical studies which use governance indicators as either left-hand or right-hand side variables are limited to small samples by the poor country coverage of many indicators. This can potentially introduce a variety of sample selection biases. In addition, the measures of precision we report can be used to correct for the attenuation bias due to measurement error in governance indicators used as dependent variables.

In the long term, however, our results also point to the inadequacy of existing governance measures. It is very unsatisfying that existing data, even with favourable assumptions, allows us to identify relatively few statistically significant differences in governance across countries. Moreover, existing data provides at best tenuous links between perceptions of governance and objective policy interventions that governments interested in improving the quality of governance can undertake. There is therefore a need to improve the quality and quantity of governance data, both by improving and extending cross-country survey work of governance perceptions, as well as employing country-specific in-depth governance diagnostics.¹⁸ Many of the polls and surveys we use suffer from deficiencies, such as poorly-worded questions about ill-defined and excessively broad concepts. There is room to improve these instruments by asking respondents about their direct experiences with well-defined events and using transparent units to measure governance. However, these are time- and resource-intensive exercises, and internationally-comparable high-quality data of this sort is years away.

¹⁸ Detailed country diagnostic exercises such as those currently being piloted by the World Bank have the potential to provide much more detailed information on the specific institutional failures which contribute to perceptions and the reality of poor governance. Kaufmann, Pradhan and Ryterman (1998) provide a description of these exercises.

References

- Efron, Bradley and Carl Morris (1971). "Limiting the Risk of Bayes and Empirical Bayes Estimators – Part 1: The Bayes Case". *Journal of the American Statistical Association*. 66:807-815.
- Efron, Bradley and Carl Morris (1972). "Limiting the Risk of Bayes and Empirical Bayes Estimators – Part 1: The Empirical Bayes Case". *Journal of the American Statistical Association*. 67:130-39.
- Goldberger, A. (1972). "Maximum Likelihood Estimation of Regressions Containing Unobservable Independent Variables". *International Economic Review*. 13:1-15.
- Hall, Robert 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.
- Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton (1999). "Governance Matters". Manuscript, The World Bank.
- Kaufmann, Daniel and Pablo Zoido-Lobaton (1999). "Corruption, Unpredictability and Performance". Manuscript, The World Bank.
- Kaufmann, Daniel, Sanjay Pradhan and Randi Ryterman (1998). "New Frontiers in Diagnosing and Combating Corruption", PREM Note Number 7, The World Bank.
- Kiefer, Nicholas M. (1980). "Estimation of Fixed Effect Models for Time Series of Cross Sections with Arbitrary Intertemporal Covariance". *Journal of Econometrics*. 14:195-202.
- Rubin, Donald (1987). *Multiple Imputation for Non-Response in Surveys*. New York: Wiley.

Table 1: Governance Indicators

Code	Name	Survey (S) or Poll (P)	Coverage	Coverage Index	Concepts Measured		
					Government Effectiveness	Rule of Law	Graft
BERI	Business Environment Risk Intelligence	P	50 mostly developed countries	0.44	Bureaucratic delays	Enforceability of contracts	"Mentality" regarding corruption
CEER	Central European Economic Review	P	26 transition economies	0.84		Rule of law	Effect of corruption on "attractiveness of country as a place to do business"
DRI	Standard and Poor's DRI	P	106 developed and developing countries	0.23	Government ineffectiveness, institutional failure	Enforceability of contracts, costs of crime	Corruption among public officials, effectiveness of anticorruption initiatives
EIU	Economist Intelligence Unit	P	115 developed and developing countries	0.19	Institutional efficacy, red tape	Crime, corruption in banking sector	Corruption among public officials
FHNT	Freedom House	P	28 transition economies	0.82	Quality of government and public administration	Rule of law	Perceptions of corruption in civil service, business interests of policymakers
GALLUP	Gallup International	S	44 mostly developed countries	0.50			Frequency of "cases of corruption" among public officials
GCS	Global Competitiveness Survey	S	59 developed and developing countries	0.42	Competence of public sector, political pressures on civil servants, time spent with bureaucrats	Citizens can file lawsuits against government, citizens accept legal adjudication, independence of judiciary, costs of crime	Frequency of "irregular payments" to officials and judiciary
GCSA	Global Competitiveness Survey, Africa	S	23 African countries	0.73	Competence of public servants, commitment to policies of previous governments	Citizens can file lawsuits against government, citizens accept legal adjudication, independence of judiciary, costs of crime	Frequency of "irregular payments" to officials and judiciary
HF	Heritage Foundation	P	160 developed and developing countries	0.06		Law and order tradition, prevalence of black market activities	
PERC	Political and Economic Risk Consultancy	S	12 Asian economies	0.83			Effect of corruption on business environment for foreign companies
PRS	Political Risk Services	P	131 developed and developing countries	0.10	Bureaucratic quality, policy stability	Rule of law	Corruption in the political system as a "threat to foreign investment"
WCY	World Competitiveness Yearbook	S	46 primarily developed countries	0.59	Efficient implementation of government decisions, political pressures on civil servants	Tax evasion, confidence in ability of authorities to protect property, confidence in administration of justice	"Improper practices" in the public sphere
WDR	World Development Report	S	74 developed and developing countries	0.25	Efficiency of government in delivering services, predictability of rules, time spent with bureaucrats	Unpredictability of the judiciary, theft and crime, ability of state to protect private property	Corruption as "obstacle to business", frequency of "additional payments" to "get things done"

Notes: Details on these sources of governance data, and definitions of the concepts measured, may be found in Appendix A of Kaufmann, Kraay and Zoido-Lobaton (1999).

Table 2: Correlations Among Governance Indicators

Government Effectiveness										
	geciu	gedri	geprs	gewdr	geberi	gefhtnt	gegcs	gegcsa	gewcy	
geciu	1.00									
	114									
gedri	0.77 *	1.00								
	96	106								
geprs	0.60 *	0.61 *	1.00							
	111	100	140							
gewdr	0.78 *	0.68 *	0.36 *	1.00						
	58	57	65	74						
geberi	0.74 *	0.71 *	0.54 *	0.73 *	1.00					
	49	50	50	30	50					
gefhtnt	0.68 *	0.71 *	0.24	0.62 *	0.67	1.00				
	19	24	21	20	6	28				
gegcs	0.76 *	0.74 *	0.52 *	0.85 *	0.70 *	0.75 *	1.00			
	64	62	72	45	46	6	75			
gegcsa	0.64 *	0.69 *	0.55 *	0.26	1.00		0.56 *	1.00		
	18	15	20	14	2	0	19	23		
gewcy	0.55 *	0.53 *	0.48 *	0.81 *	0.65 *	0.06	0.92 *		1.00	
	43	44	46	27	41	4	46	0	46	

Rule of Law													
	rl dri	rl eiu	rl hf	rl prs	rl wrdr	rl beri	rl ceer	rl fhnt	rl gcs	rl gcsa	rl jkz	rl wcy	rl score
rl dri	1.00												
	106												
rl eiu	0.73 *	1.00											
	96	114											
rl hf	0.73 *	0.68 *	1.00										
	105	112	160										
rl prs	0.75 *	0.75 *	0.62 *	1.00									
	100	111	137	140									
rl wrdr	0.58 *	0.75 *	0.76 *	0.59 *	1.00								
	57	58	72	65	74								
rl beri	0.73 *	0.70 *	0.82 *	0.62 *	0.54 *	1.00							
	50	49	50	50	30	50							
rl ceer	0.89 *	0.79 *	0.90 *	0.52 *	0.51 *	0.66	1.00						
	24	19	25	20	20	6	27						
rl fhnt	0.76 *	0.68 *	0.86 *	0.33	0.39 *	0.75 *	0.91 *	1.00					
	24	19	26	21	20	6	27	28					
rl gcs	0.70 *	0.78 *	0.78 *	0.70 *	0.74 *	0.77 *	0.96 *	0.78 *	1.00				
	53	56	59	59	33	46	6	6	59				
rl gcsa	0.43	0.31	0.03	0.27	0.61 *	-1.00			0.87	1.00			
	15	18	23	20	14	2	0	0	3	23			
rl jkz	0.63 *	0.45 *	0.37 *	0.57 *	0.13	0.64 *	0.19	0.00	0.50 *	0.16	1.00		
	69	73	76	74	48	44	19	20	49	8	77		
rl wcy	0.63 *	0.73 *	0.77 *	0.68 *	0.78 *	0.75 *	0.98 *	0.94 *	0.94 *		0.47 *	1.00	
	44	43	46	46	27	41	4	4	46	0	40	46	
rl score	1.00 *	0.99 *	0.96 *	0.91 *	0.72	0.63	0.99 *	0.98 *	0.96 *		0.37	0.96 *	1.00
	5	5	5	5	5	5	5	5	5	0	5	4	5

Graft													
	gr dri	gr eiu	gr prs	gr wrdr	gr beri	gr ceer	gr fhnt	gr gallup	gr gcs	gr gcsa	gr perc	gr wcy	gr score
gr dri	1.00												
	106												
gr eiu	0.80 *	1.00											
	97	115											
gr prs	0.65 *	0.64 *	1.00										
	100	112	140										
gr wrdr	0.69 *	0.80 *	0.46 *	1.00									
	57	58	65	74									
gr beri	0.57 *	0.58 *	0.46 *	0.78 *	1.00								
	50	50	50	30	50								
gr ceer	0.91 *	0.76 *	0.68 *	0.56 *	0.38	1.00							
	24	19	20	20	6	26							
gr fhnt	0.79 *	0.60 *	0.72 *	0.58 *	0.40	0.92 *	1.00						
	23	19	21	20	6	25	28						
gr gallup	0.63 *	0.78 *	0.62 *	0.81 *	0.46 *	0.71 *	0.59 *	1.00					
	42	42	44	26	29	9	9	44					
gr gcs	0.77 *	0.88 *	0.57 *	0.87 *	0.73 *	0.98 *	0.80 *	0.80 *	1.00				
	53	57	59	33	46	6	7	35	59				
gr gcsa	0.59 *	0.53 *	0.45 *	0.61 *	1.00			0.95	0.79	1.00			
	15	18	20	14	2	0	0	3	3	23			
gr perc	0.55 *	0.95 *	0.33	0.96 *	0.76 *			0.84 *	0.89 *		1.00		
	12	12	12	6	11	0	0	5	12	0	12		
gr wcy	0.69 *	0.84 *	0.65 *	0.93 *	0.62 *	-0.32	-0.51	0.87 *	0.86 *		0.93 *	1.00	
	44	44	46	27	41	4	4	31	46	0	11	46	
gr score	0.91 *	0.92 *	0.82 *	0.86 *	0.74 *	0.89 *	0.87 *	0.73 *	0.82 *	0.64 *	0.82 *	0.82 *	1.00
	51	51	51	51	30	13	13	25	33	11	6	27	51

This table reports pairwise correlations between governance indicators within each governance cluster. The numbers below the correlation coefficients indicate the number of countries common to each pair of indicators. * indicates significance at the 90% level.

Table 3: Parameter Estimates

	Government Effectiveness			Rule of Law			Graft		
	$\alpha(k)$	$\beta(k)$	$\sigma\varepsilon(k)$	$\alpha(k)$	$\beta(k)$	$\sigma\varepsilon(k)$	$\alpha(k)$	$\beta(k)$	$\sigma\varepsilon(k)$
Representative Indicators									
DRI	0.539	0.239	0.588	0.668	0.179	0.583	0.539	0.221	0.618
EIU	0.432	0.226	0.396	0.442	0.285	0.445	0.312	0.309	0.322
HF				0.466	0.247	0.751			
PRS	0.789	0.109	1.222	0.606	0.220	0.656	0.506	0.142	1.129
WDR	0.473	0.097	0.685	0.354	0.135	0.681	0.465	0.150	0.636
Non-Representative Indicators									
BERI	0.406	0.133	0.707	0.404	0.136	0.687	0.483	0.173	1.226
CEER				0.604	0.380	0.303	0.615	0.359	0.356
FHNT	0.624	0.396	0.440	0.566	0.397	0.454	0.561	0.513	0.511
GALLUP							0.470	0.149	0.709
GCS	0.398	0.151	0.600	0.526	0.148	0.594	0.493	0.247	0.457
GCSA	0.470	0.112	0.450	0.499	0.041	1.695	0.498	0.297	0.406
PERC							0.302	0.299	0.270
WCY	0.307	0.120	0.906	0.382	0.156	0.580	0.212	0.284	0.498

Table 4: Assigning Countries to Quartiles

Proportion of Countries for Which an x% Confidence Interval Lies Entirely in the Indicated Quartile

	<u>x=90%</u>	<u>x=75%</u>	<u>x=50%</u>
Government Effectiveness			
First Quartile	0.31	0.54	0.72
Second Quartile	0.00	0.00	0.26
Third Quartile	0.00	0.13	0.31
Fourth Quartile	0.59	0.69	0.79
Rule of Law			
First Quartile	0.31	0.43	0.65
Second Quartile	0.00	0.05	0.39
Third Quartile	0.12	0.24	0.55
Fourth Quartile	0.55	0.63	0.84
Graft			
First Quartile	0.13	0.23	0.49
Second Quartile	0.00	0.03	0.26
Third Quartile	0.08	0.21	0.36
Fourth Quartile	0.65	0.72	0.80

This table reports the fraction of all countries whose point estimate of governance falls in the indicated quartile for which the corresponding x% confidence interval also falls entirely within that quartile, for each of the three governance aggregates and for a range of values of x.

Table 5: Consequences of Correlated Disturbances

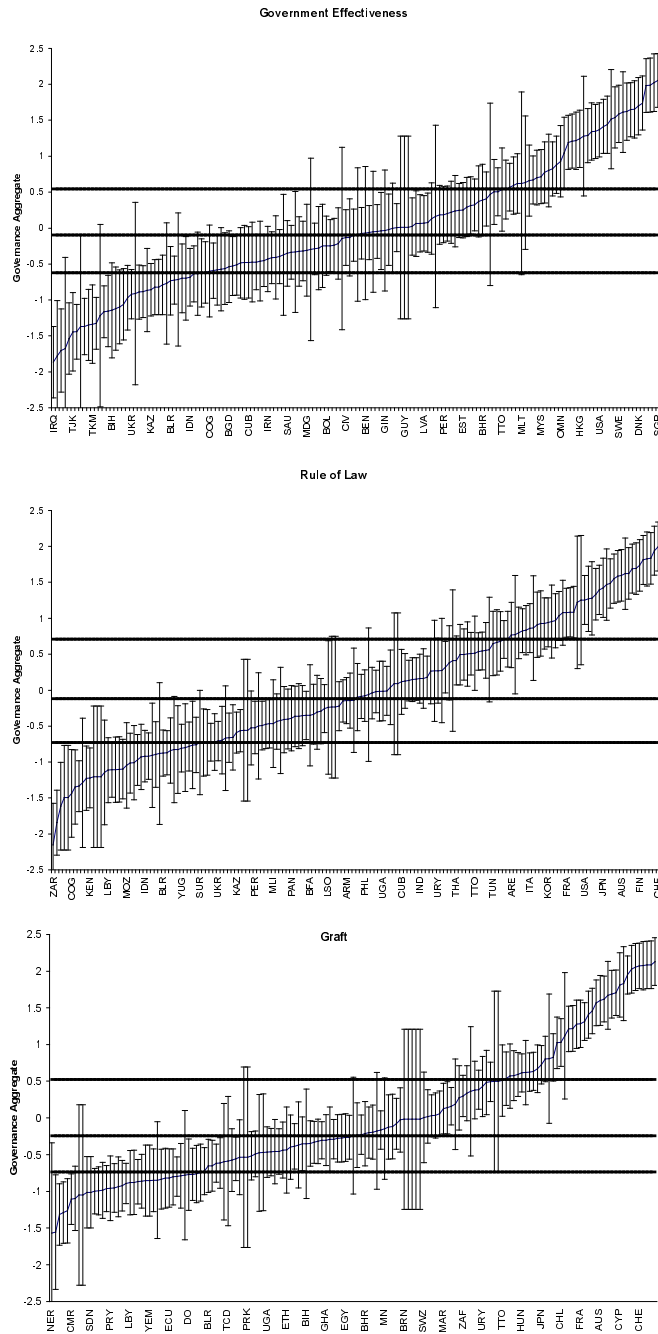
Average Standard Error of Governance Aggregate
Based on Representative Indicators

Assumed Error Correlation:
 $\rho = 0$ $\rho = 0.25$ $\rho = 0.5$

Government Effectiveness	0.32	0.31	0.35
Rule of Law	0.33	0.47	0.66
Graft	0.31	0.42	0.44

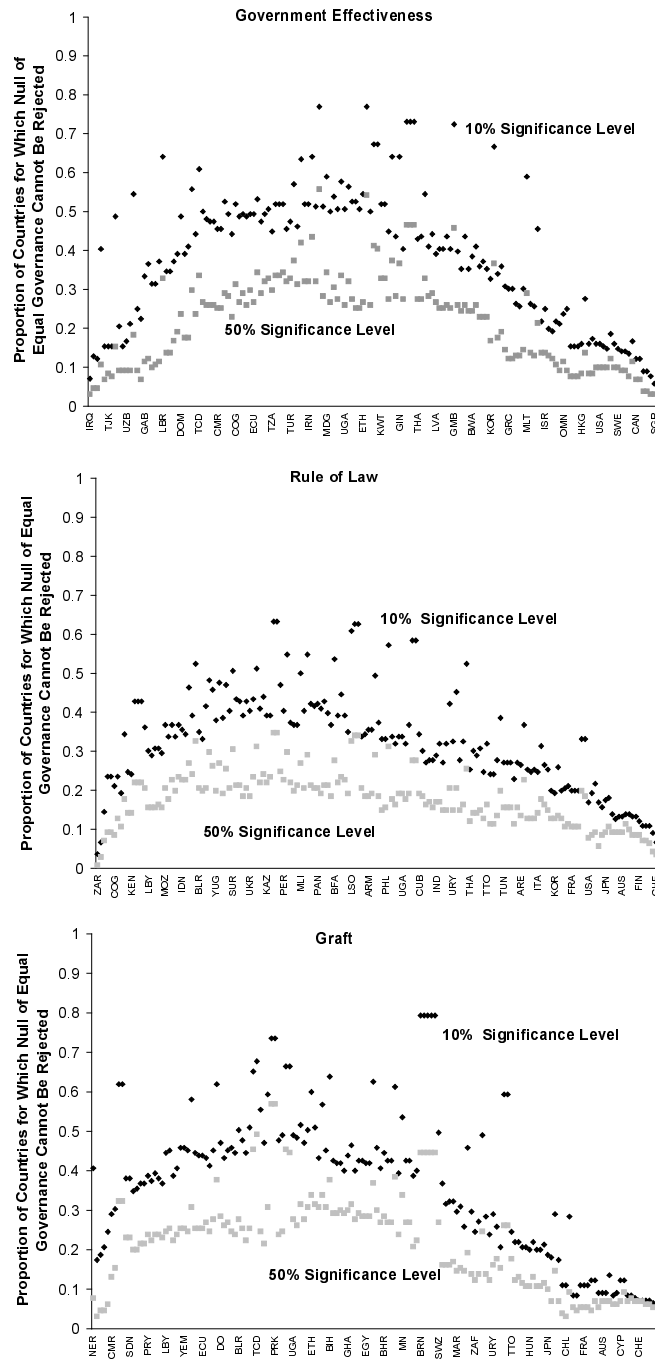
This table reports the standard error of an aggregate indicator (based on a balanced panel of three sources), under alternative assumptions regarding the correlation of the disturbances.

Figure 1 : Estimates of Governance



Countries are ordered on the horizontal axis in ascending order according to their point estimates of governance, and their point estimates and 90% confidence intervals are indicated on the vertical axis. The horizontal lines delineate the quartiles of the distribution of governance estimates. For reasons of space, country names are indicated only for every fifth country.

Figure 2 : Significance of Pairwise Governance Comparisons



Countries are ordered on the horizontal axis in ascending order according to their point estimates of governance. For each country, we plot on the vertical axis the proportion of all comparator countries for which the probability that governance in the reference country is greater than that in the comparator country is either between 5% and 95% (dark dots), or is between 25% and 75% (light dots). For reasons of space, country names are indicated only for every fifth country.

**Figure 3 : Precision of Governance Aggregates
as a Function of Pairwise Correlations of Observed Data**

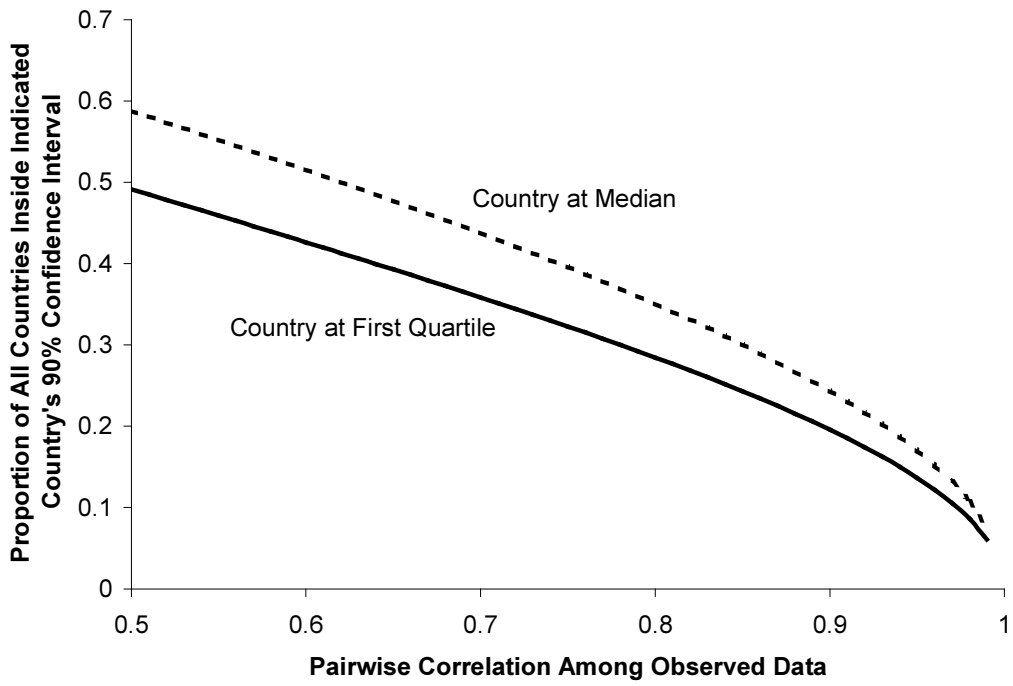
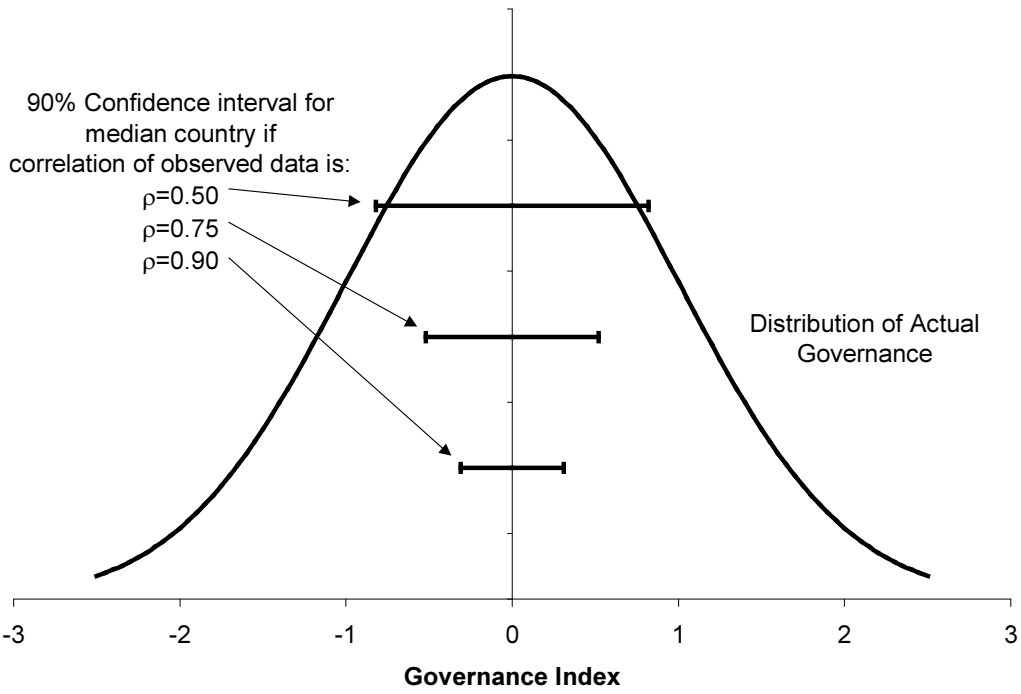
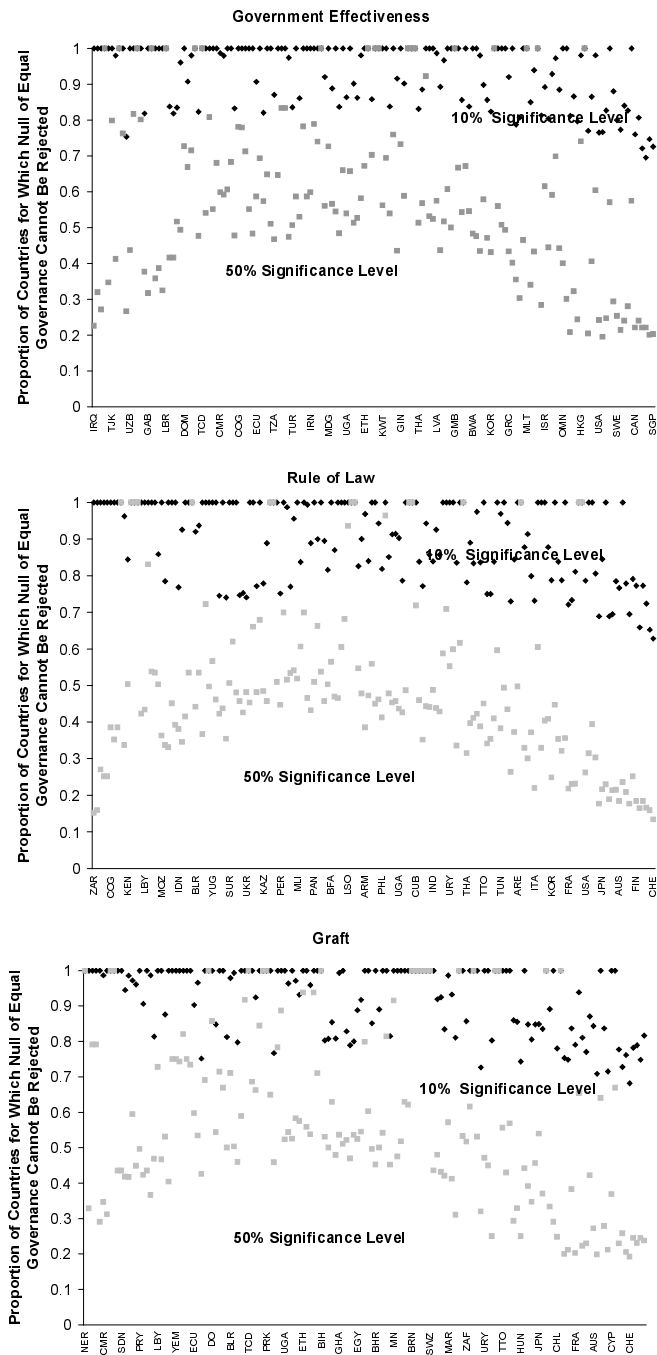
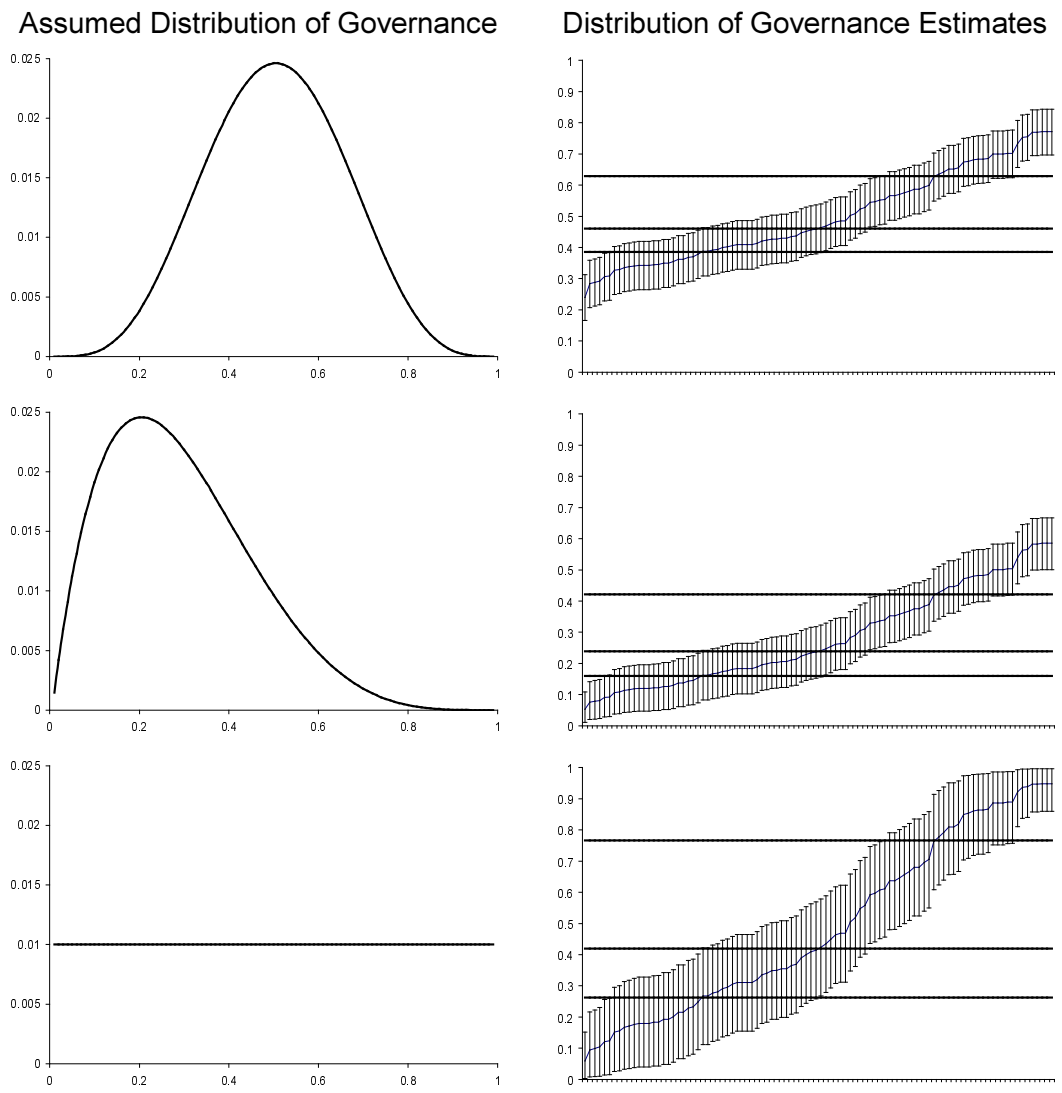


Figure 4 : Significance of Pairwise Governance Comparisons Based on Ordinal Method



Countries are ordered on the horizontal axis in ascending order according to their point estimates of governance. For each country, we plot on the vertical axis the proportion of all comparator countries for which the probability that governance in the reference country is greater than that in the comparator country is either between 5% and 95% (dark dots), or is between 25% and 75%. For reasons of space, country names are indicated only for every fifth country.

Figure 5 : Alternative Distributions for Governance



The rows of this figure correspond to the assumptions that unobserved governance follows a Beta(5,5), Beta(2,5) and a Beta(1,1) distribution. In the left column we show the shape of this distribution. In the right column, we show the corresponding analog to Figure 1, for a governance aggregate constructed using a balanced panel of three indicators of graft.