

Governance Matters

Main Text

Daniel Kaufmann
Aart Kraay
Pablo Zoido-Lobaton

Six new aggregate measures capturing various dimensions of governance provide new evidence of a strong causal relationship from better governance to better development outcomes.



Summary findings

In a cross-section of more than 150 countries, Kaufmann, Kraay, and Zoido-Lobaton provide new empirical evidence of a strong causal relationship from better governance to better development outcomes. They base their analysis on a new database containing more than 300 governance indicators compiled from a variety of sources. They provide a detailed description of each of these indicators and sources. Using an unobserved components methodology (described in

the companion paper by Kaufmann, Kraay, and Zoido-Lobaton, “Aggregating Governance Indicators,” Policy Research Working Paper 2195), they then construct six aggregate indicators corresponding to six basic governance concepts: voice and accountability, political instability and violence, government effectiveness, regulatory burden, rule of law, and graft. As measured by these indicators, governance matters for development outcomes.

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. (60 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.

Governance Matters

Daniel Kaufmann
Aart Kraay
Pablo Zoido-Lobaton

The World Bank

October 1999

MAIN TEXT ONLY

Appendices for this paper can be found at
http://www.worldbank.org/wbi/governance/gov_pdfs/govmatrs-annex.pdf

A companion paper, "Aggregating Governance Indicators" can be found at
http://www.worldbank.org/wbi/governance/gov_pdfs/agg_ind.pdf

Abstract: Recent interest in the consequences of governance for development outcomes has been accompanied by a proliferation of data measuring various aspects of governance. In this paper, we describe a new database containing over 300 governance indicators compiled from a variety of sources. We provide a detailed description of each of these indicators and sources, and construct six aggregate indicators corresponding to six basic governance concepts. Governance, as measured by these indicators, matters. In a cross-section of over 150 countries, we provide new empirical evidence of a strong causal relationship from better governance to better development outcomes.

The World Bank, 1818 H Street N.W., Washington, D.C. 20433. (dkaufmann@worldbank.org, akraay@worldbank.org, pzoidolobaton@worldbank.org). The opinions expressed here are the authors', and in no way reflect those of the World Bank, its Executive Directors, or the countries they represent. We would like to thank Serhiy Filipov for diligent research assistance.

1. Introduction

In recent years there has been a surge of interest in the consequences of governance and misgovernance for development. This has been accompanied by a proliferation of data measuring subjective perceptions of various aspects of governance. In this paper, we describe a new governance database containing over 300 such governance measures compiled from a variety of sources.¹ We provide a detailed description of each of these indicators and sources, and construct six aggregate indicators corresponding to six fundamental governance concepts. We conclude with new empirical evidence that governance matters, in the sense that there is a strong causal relationship from good governance to better development outcomes such as higher per capita incomes, lower infant mortality, and higher literacy.

In order to organize our data collection and dissemination efforts, we require a working definition of governance itself.² We define governance broadly as the traditions and institutions by which authority in a country is exercised. This includes (1) the process by which governments are selected, monitored and replaced, (2) the capacity of the government to effectively formulate and implement sound policies, and (3) the respect of citizens and the state for the institutions that govern economic and social interactions among them.

The focus of our data collection effort is a set of indicators which measure subjective perceptions regarding the quality of governance in different countries. Our data are drawn from two types of sources: polls of experts, which reflect country ratings produced by commercial risk rating agencies and other organizations, and cross-country surveys of residents carried out by international organizations and other non-governmental organizations. We do not attempt to compile or present the wide array of available quantitative and descriptive data on cross-country differences in political and social institutions.³ While these are certainly important determinants of the cross-

¹ Pending appropriate authorization from original sources, we plan to have the database itself ready for dissemination on the Web late in 1999.

² There does not appear to be a single accepted definition of governance. Our definition of governance is in part motivated those suggested by the Institute for Governance, IDEA, and the IMF.

³ A database of such objective institutional indicators is currently being compiled as a joint effort by Phil Keefer (World Bank), Robert Bates (Harvard University), and David Epstein and Sharyn O'Halloran

country differences in the quality of governance, our focus here is limited to measuring the perceptions of interested parties -- residents of a country, entrepreneurs, foreign investors, and civil society at large -- regarding the quality of governance in a country. Although this kind of data is inherently subjective, there are several reasons why it is useful in measuring governance. First, for many issues such as the prevalence of corruption, objective data is almost by definition rather difficult to obtain. There are few alternatives to subjective indicators if one wishes to measure these aspects of governance. Second, perceptions of the quality of governance may often be as important as objective differences in institutions across countries. While a country may nominally enjoy a set of sound institutions according to certain standards, the confidence of residents of a country in these institutions is also required if they are to contribute to good governance. Third, we have found in other work that subjective perceptions can have significant explanatory power for future economic outcomes.⁴

We use our definition of governance to organize a subset of the governance indicators into six clusters. We summarize two key aspects of the process by which those in authority are selected and replaced with clusters labelled "Voice and Accountability", and "Political Instability and Violence". We capture the capacity of the state to implement sound policies with two clusters we refer to as "Government Effectiveness" and "Regulatory Burden". Finally, two clusters labelled "Rule of Law" and "Graft" capture the respect of citizens and the state for the rules which govern their interactions. For each of these six clusters, we use the procedure laid out in Kaufmann, Kraay and Zoido-Lobaton (1999) to construct aggregate governance indicators. The advantage of these aggregate indicators is that they span a much larger set of countries than most individual indicators, permitting cross-country comparisons of between 155 and 173 countries, depending on the aggregate. More important, the aggregation method we use provides quantitative measures of the precision of both the aggregate governance indicators and their components, allowing formal statistical testing of hypotheses regarding cross-country differences in governance along these dimensions.

(Stanford University), in an effort coordinated by the Center for International Development at Harvard University.

⁴ In the context of the East Asian financial crisis, Kaufmann, Mehrez and Schumkler (1999) find that investor perceptions of future financial instability had significant explanatory power for future actual volatility, over and above available macroeconomic data.

There is a growing empirical literature documenting the relationship between indicators of various aspects of governance and economic outcomes.⁵ Using the six aggregate governance indicators described above, we contribute to this literature by providing new empirical evidence of a strong causal relationship from improved governance to better development outcomes. Following the approach of Hall and Jones (1999), we estimate a series of very parsimonious regressions of the log-level of per capita income on each of the six aggregate governance indicators in turn. We correct for reverse causation, omitted variable bias, and measurement error in the governance variables, by using the shares of the population speaking European languages as instruments.⁶ We find that governance matters a great deal for economic outcomes. In particular, a one-standard deviation increase in any one of our governance indicators causes between a two-and-a-half and four-fold increase (decrease) in per capita incomes (infant mortality), and a 15 to 25 percent increase in literacy. Tests of overidentifying restrictions and summary statistics of the explanatory power of the first-stage regressions provide support for the validity of this modelling strategy.

The remainder of this paper proceeds as follows. Section 2 briefly describes the governance database, with the details of each source of data and each governance measure relegated to a lengthy Appendix 1. Section 3 summarizes the components of each of the six governance aggregates and explains the aggregation procedure. The details of the composition of each aggregate are presented in a series of tables in Appendix 2. Section 4 contains new empirical evidence on the relationship between governance and development outcomes.

⁵ This literature includes Mauro (1995) on the effects of corruption on economic growth and investment; Loayza (1996) on the determinants of the unofficial economy; Ades and DiTella (1996) on the causes and consequences of corruption; Knack and Keefer (1997) on the importance of institutions for economic growth; Tanzi and Davoodi (1997) on corruption and public investment; Wei (1997) on the effects of corruption on FDI; Rodrik (1997) on the role of institutions in the success of East Asia ; Johnson, Kaufmann and Zoido-Lobaton (1998) on the effects of corruption on the unofficial economy; Hall and Jones (1999), on the relationship between levels of per capita income and a measure of what they call “social infrastructure”, and Chong and Calderón (1999) on the Granger-causality between institutions and economic growth.

⁶ Hall and Jones (1999) provide a detailed justification of such instruments. Recent evidence such as that presented by La Porta, López-de-Silanes, Shleifer and Vishny (1999) that legal origin, ethnic fractionalization and other relatively exogenous factors predict cross-country differences in the quality of government provides further support for this identification strategy.

2. The Governance Database

We have compiled a large number of governance measures from a variety of sources into a governance database. These different sources produce indicators of a wide range of concepts relating to governance, using different techniques, and covering different sets of countries. Table 1 presents the sources of governance data included in the database. These include international organizations, political and business risk rating agencies, think tanks, and non-governmental organizations. In this section, we provide an overview of some of the key features of these sources. Appendix 1 presents a detailed description of each of these sources.

We classify sources of governance data along two dimensions: (1) according to their “nature”, i.e. the techniques they employ to measure governance; and (2) according to their “coverage”, i.e. the extent to which the set of countries covered by each indicator is representative of the world as a whole.

With respect to their “nature”, sources are either polls of experts or cross-country surveys of firm managers or citizens in general. Measures based on polls of experts represent consensus ratings agreed upon by a small number of country, sector, and regional experts convened by the reporting organization. Typically, country analysts produce an initial assessment for each country based on publicly-available information and their direct knowledge of the country. For most of the sources we report, these assessments are guided by a checklist of specific issues which analysts take into consideration when providing their initial ratings. For all sources where we were able to obtain this information, we report the checklists underlying the ratings. These initial ratings are then reviewed by a panel of regional and sectoral experts, who determine the final rating for each country. In addition to providing quality control on the initial assessments, the purpose of this final stage is to improve the cross-country comparability of the ratings by ensuring that countries are benchmarked appropriately.

Indicators based on surveys present averages by country of the responses of a large number of respondents to a variety of questions relating to governance. Typically, survey respondents are asked to rate aspects of governance on a categorical scale. The sampling frame of the surveys we use varies. Some sources focus on the opinions of

the business community, others focus on the opinions of expatriates, and some are broad-based surveys of citizens.

The two types of sources of governance data each have their advantages and disadvantages. The main advantage of polls of experts is that they are explicitly designed for cross-country comparability, and considerable effort is placed in the benchmarking process which ensures this. However, the difficulties with such measures are also clear. Typically they are based on the opinions of only a few experts per country, and the quality of the country ratings depends to a great extent on the knowledge of experts regarding the countries they are assessing. The ratings are particularly prone to two types of feedback: countries with good economic outcomes may be more likely to receive favourable ratings, and country rankings by other organizations are frequently an input into the rating process of each organization. Country ratings may also be affected by the political or ideological agenda of the organization producing the ratings. Despite these difficulties, we believe that there are at least two reasons to think that on average, such sources provide valuable information on governance. First, we find that the indicators produced by such organizations generally correlate strongly with measures based on surveys of residents and entrepreneurs. Second, we find the fact that commercial rating organizations such as EIU, DRI and PRS are able to consistently sell their assessments to commercial subscribers for considerable fees to be convincing evidence that these sources are in fact producing useful information.⁷

The advantage of surveys is that they reflect the opinions of a larger number of respondents that are more closely connected with the countries they are assessing. Nevertheless, they suffer from two disadvantages. First, survey questions can be interpreted in context- or culture-specific ways. For example, a response regarding the prevalence of “improper practices” is coloured by country-specific perceptions of what “improper practices” are perceived to be. This will hinder the cross-country comparability of responses to otherwise identical questions.⁸ Second, cross-country

⁷ For example, as of May 1999 an annual subscription to EIU’s Country Risk Service costs \$625 US per year, an annual subscription to PRS’s International Country Risk Guide on CD-ROM costs \$4500 US per year.

⁸ A related problem is that residents of particular countries may have a broad-based predisposition to overstate or understate governance problems in that country (see Kaufmann and Zoido-Lobaton (1999) and Kaufmann and Wei (1999)).

surveys relating to governance are very costly to design and implement, and as a result typically cover a much smaller set of countries than polls of experts.

Sources of governance data also vary with respect to the sample of countries they cover. A number of sources cover a very large sample of developed and developing countries (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 sources cover primarily developed countries but also include major developing countries (WCR, GALLUP, BERI). Since there is a strong positive association across countries between governance and per capita incomes, this difference between sources makes it difficult to compare indicators from sources which cover sets of countries with very different income levels. Similarly, there may be regional differences in governance which hamper comparisons across sources. For example, it is not clear how to compare a governance rating based only on transition economies with one based on a broad set of countries. As discussed in our companion paper (Kaufmann, Kraay and Zoido-Lobatón (1999)), the methodology we use to construct aggregate governance indicators takes these differences in country coverage into account. Users of the governance database are also encouraged to keep this consideration in mind when comparing information from different sources.

In order to distinguish between indicators according to the representativeness of their country coverage, we construct a simple index which measures differences between the distribution of countries across income and regional classifications and the corresponding distribution of all countries in the world. 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 values of the deviations between the share of countries in each of the 45 categories (five income categories x nine regions) 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 six indicators covering the largest number of countries (DRI, EIU, FH, HFWSJ, PRS and WDR) are substantially more representative according to this measure than the others, with a value of this index of less than 0.25. We refer to these as representative indicators, and the remainder as non-representative indicators.

3. Aggregate Governance Indicators

In this section, we organize a subset of the governance indicators into six clusters corresponding to six basic aspects of governance, and describe how these indicators can be combined into aggregate governance indicators. In Appendix 2, we present six tables which identify the specific components of each governance cluster with reference to the tables describing sources in Appendix 1. The aggregate governance indicators described here are based on data for 1997 and 1998.⁹ We stress at the outset that the classification of indicators into clusters is not intended to be definitive. Rather, it reflects our own views of what constitutes a useful and interesting organization of the data that is consistent with prevailing notions of governance.¹⁰

Governance Clusters

The first two governance clusters are intended to capture the first part of our definition of governance: the process by which those in authority are selected and replaced. We refer to the first of these as “Voice and Accountability”, and include in it a number of indicators measuring various aspects of the political process, civil liberties and political rights. These indicators measure the extent to which citizens of a country are able to participate in the selection of governments. We also include in this category three indicators measuring the independence of the media, which serves an important role in holding monitoring those in authority and holding them accountable for their actions.

The second governance cluster is labelled “Political Instability and Violence”. In this index we combine several indicators which measure perceptions of the likelihood that the government in power will be destabilized or overthrown by possibly unconstitutional and/or violent means. This index captures the idea that the quality of governance in a country is compromised by the likelihood of wrenching changes in

⁹ As we expand the governance database, we hope to construct comparable aggregates for other time periods, enabling comparisons over time as well as across countries.

¹⁰ Users of the governance database may question both the classification of individual indicators into these six clusters, and also the six categories themselves. Users are therefore welcomed to use individual indicators and construct aggregate indicators that are most useful to their own analytic purposes

government, which not only has a direct effect on the continuity of policies, but also at a deeper level undermines the ability of all citizens to peacefully select and replace those in power.

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

The last two clusters summarize in broad terms the respect of citizens and the state for the institutions which govern their interactions. In “rule of law” we include several indicators which measure the extent to which agents have confidence in and abide by the rules of society. These include perceptions of the incidence of both violent and non-violent crime, the effectiveness and predictability of the judiciary, and the enforceability of contracts. Together, these indicators measure the success of a society in developing an environment in which fair and predictable rules form the basis for economic and social interactions. The final cluster, which we refer to as “graft”, measures perceptions of corruption. Unlike all of the other clusters, this one can be easily described using the conventional definition of corruption: the exercise of public power for private gain. Despite this straightforward focus, the particular aspect of corruption measured by the various sources differs somewhat, ranging from the frequency of “additional payments to get things done” to the effects of corruption on the business environment. The presence of corruption is often a manifestation of a lack of respect of both the corrupter (typically a private citizen) and the corrupted (typically a public official) for the rules which govern their interactions, and hence represents a failure of governance according to our definition.

Aggregating Governance Indicators

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

For each of these clusters, we combine the component indicators into an aggregate governance indicator using the method described in Kaufmann, Kraay and Zoido-Lobaton (1999). We use an unobserved components model which expresses the observed data in each cluster as a linear function of the unobserved common component of governance, plus a disturbance term capturing perception errors and/or sampling variation in each indicator.¹¹ This model enables us to compute estimates of each of the six governance measures for each country, as well as measures of the precision of these estimates. Formally, the estimate of governance for a country produced by the unobserved components model is the mean of the distribution of unobserved governance conditional on the observed data for that country. This conditional mean is simply a weighted average of appropriately-rescaled scores of each of the component indicators. We also report the standard deviation of this conditional distribution as an indicator of the confidence we can have in this estimate. This standard deviation is declining in the number of individual indicators in which a particular country appears, and is increasing in the variance of the disturbance term on each of these indicators. Our choice of units for governance ensures that the estimates of governance have a mean of zero, a standard deviation of one, and range from around -2.5 to around

¹¹ If there are multiple questions on a single source which belong in a particular cluster, we first average questions from the same source, and treat the average as a single indicator based on that source. We do this because the necessary assumption that the errors on indicators are independent across indicators is much less tenable in the case of multiple questions from the same source.

2.5.¹² We orient these aggregate indicators such that higher values correspond to better outcomes.

The assumptions of the unobserved components model ensure that the distribution of governance in each country is normal, conditional on the data for that country. Therefore, these conditional means and standard deviations for each country have a natural interpretation. For example, a useful interpretation of the reported estimates and standard deviations for each country is to note that there is a 90% probability that the “true” level of governance in a country is in an interval of plus or minus 1.64 times the reported standard deviation centered on the point estimate itself. We refer to such a range as a 90% confidence interval around the estimate of governance for a country.¹³ Similarly, since the assumptions of the model ensure that the distribution of governance conditional on the observed data is independent across countries, it is possible to make probabilistic statements comparing governance in pairs of countries. For example, it is straightforward to compute the probability that governance in one country j is better than in another country j' by evaluating the appropriate cumulative bivariate normal distribution.

As emphasized in our paper on aggregating governance indicators, we find that the underlying governance concepts in each cluster are not very precisely estimated, in the sense that the measured standard deviations are large relative to the units in which governance is measured. We illustrate this point in Figure 1. In each of the six panels, we order countries in ascending order according to their point estimates of governance on the horizontal axis, and on the vertical axis we plot the estimate of governance and the associated 90% confidence interval described above. The size of these confidence intervals varies across countries, as different countries appear in different numbers of sources with different variances. The resulting confidence intervals are large relative to the units in which governance is measured. To emphasize this key point, the horizontal lines in Figure 1 delineate the quartiles of the distribution of governance estimates. Relatively few countries have 90% confidence intervals that lie entirely within a given

¹² Underlying the model is the assumption that the distribution of governance is normal with zero mean and unit standard deviation. The distribution of the actual estimates of governance is very similar.

¹³ This is a slight abuse of terminology, as these are not confidence intervals in the usual frequentist sense of a stochastically varying interval centered around a fixed unknown parameter. Rather, we treat governance as a random variable, and the 90% confidence interval is simply the 5th and 95th percentiles of the conditional distribution of governance given the observed data.

quartile. This indicates that sorting countries into even broad categories is subject to significant margins of error for most countries.

The rather large size of these confidence intervals has important implications for the use of these aggregate governance indicators. It is clear that small differences in point estimates of governance across countries are not likely to be statistically significant. This mirrors the reality that it is in fact difficult to distinguish small differences among countries using this type of data on governance. As a result, users of this data should focus on the range of possible governance for each country as summarized in the 90% confidence intervals shown in Figure 1. For two countries at opposite ends of the scale of governance, whose 90% confidence intervals do not overlap, it is reasonable to conclude that there are in fact significant differences in governance between these two countries. For pairs of countries that are closer together and whose 90% confidence intervals overlap, one should be much more circumspect about the significance of estimated differences in governance between two such countries.

Despite the imprecision of these aggregate indicators, they are still very useful, for several reasons. First, since each of these aggregate indicators spans a much larger set of countries than any individual indicator, it is possible to make comparisons – however imprecise -- across a much larger set of countries than would be possible with any single indicator. Second, although imprecise, each aggregate indicator provides a more precise signal of its corresponding broader governance concept than do any of its component indicators, and moreover provides a convenient and consistent summary of the available evidence. Third, the measures of precision for each country are useful because they enable formal statistical tests of cross-country differences in governance instead of arbitrary comparisons. Fourth, it is possible to use information in the estimates of the precision of each aggregate to quantify the effect of measurement error in regression analyses that use governance indicators as right-hand side variables.

4. Governance Matters

How much does governance matter for development outcomes? In Figures 2-4, we document the strong positive association between each of the six aggregate governance indicators and three development outcomes: per capita incomes, infant mortality, and adult literacy. However, these simple correlations cannot be interpreted as evidence of a causal relationship from better governance to better development outcomes, since they do not control for other determinants of development, and may also reflect a reverse causation from development to governance. In this section, we disentangle the direct effects of governance on development outcomes by instrumenting for governance in a series of cross-sectional regressions of these outcomes on each of our six governance aggregates. We find that improvements in governance have a very large payoff in terms of development outcomes.

Our approach is motivated by Hall and Jones (1999), who argue that the key determinant of cross-country differences in output per worker is differences in what they call “social infrastructure”, and which can be broadly interpreted as some combination of the aspects of governance we discuss above. They measure social infrastructure as a simple average of a number of governance indicators from PRS and a variable measuring trade openness constructed by Sachs and Warner (1995). We extend these results by considering a wider range of governance indicators and development outcomes in a larger sample of countries.

Empirical Specification

To simplify the discussion, we present the empirical specification using the logarithm of per capita GDP as the dependent variable. When we turn to the results, we will consider the identical specifications for the two other development indicators: infant mortality and adult literacy. We follow Hall and Jones (1999) in writing an extremely parsimonious specification in which the logarithm of per capita income, y_j , is a linear function of governance, g_j , and an error term e_j :

$$(1) \quad y_j = \alpha + \beta \cdot g_j + e_j$$

The error term reflects both measurement error in per capita incomes as well as all other determinants of cross-country differences in per capita incomes that have been excluded from this simple specification. To the extent that these are correlated with governance, this will introduce omitted variables bias in simple ordinary least squares (OLS) estimates of Equation (1).

Governance itself is not randomly distributed across countries. Good governance requires time and resources to develop, suggesting that richer countries are more likely to enjoy good governance. Governance also depends on a country's political and social history, especially in those countries that inherited a set of institutions from the former colonial powers. Based on this motivation we write governance as a function per capita income y_j , as well as a set of additional observable determinants x_j :

$$(2) \quad g_j = \gamma + \delta \cdot y_j + \theta' x_j + v_j$$

where v_j is a zero-mean error term capturing other unobserved determinants of governance. Depending on the sign of the feedback from income to governance, simple OLS estimates of (1) will be biased upwards or downwards.

Finally, we have already observed that governance itself is not very precisely measured. In particular, our observed governance aggregates, g_j^* , provide a noisy signal of "true" governance:

$$(3) \quad g_j^* = g_j + u_j$$

where u_j is a zero-mean disturbance term reflecting measurement error. This will introduce a downward bias in regressions using observed governance as the explanatory variable.

Substituting Equation (3) into Equation (1) yields our empirical specification, in which we regress the log-level of per capita income, y_j , on observed governance, g_j^* :

$$(4) \quad y_j = \alpha + \beta \cdot g_j^* + (e_j - \beta \cdot u_j)$$

We identify the model with the assumption is that the observable determinants of governance included in x_j are uncorrelated with the composite error term in Equation (4). In particular, these determinants of governance must be uncorrelated with both the other determinants of income excluded from our simple specification (captured by e_j), and with the measurement error in governance (captured by u_j), i.e. $E[x_j \cdot e_j] = 0$ and $E[x_j \cdot u_j] = 0$. Under this assumption, the elements of x_j are valid instruments for measured governance g_j^* in Equation (4). Given such instruments, we can address the problems of omitted variables bias, simultaneity bias, and attenuation bias using two-stage least squares (2SLS).

We follow Hall and Jones (1999) in using the fraction of the population speaking English, and the fraction of the population speaking a major European language (English, French, German, Portuguese or Spanish) as instruments. As Hall and Jones (1999) argue, a major feature of world history in the past several centuries has been the spread of Western European influence across countries. Together with this influence came an institutional and cultural background that in certain countries was conducive to the development of strong institutions and good governance. Since the extent of this influence varies widely across countries and may safely be thought of as exogenous to economic outcomes centuries later, proxies for this influence such as the linguistic variables are good candidates for instruments.¹⁴

Data

Our data are drawn from readily-available sources. The six governance indicators have already been discussed in the previous sections, and refer to the period 1997-98. We measure per capita income as the average over all available years

¹⁴ Hall and Jones (1999) also use distance from the equator as a proxy for the strength of past Western European influences. We found this not to be a good instrument in our regressions as we could not reject the null hypothesis that it has a direct influence on per capita incomes in addition to its indirect effect through governance. Only if we combine our direct measures of governance with other measures of policy (such as the Sachs-Warner index used by Hall and Jones (1999)) does an instrument set including distance from the equator pass tests of overidentifying restrictions. Similarly, in the Hall and Jones (1999) specification, tests of overidentifying restrictions fail if the policy variable is excluded from their measure of social infrastructure.

between 1990 and 1995 of per capita GDP in 1987 US dollars adjusted for purchasing power parity.¹⁵ For 152 countries, we obtain this data directly from World Bank estimates of PPP-adjusted per capita GDP as reported in the World Development Indicators. For a further 10 countries we have data on per capita GDP converted to dollars using market exchange rates in the World Development Indicators, and we use a regression of the log-level of per capita GDP at PPP on the log-level of per capita GDP at market exchange rates to crudely impute PPP adjustments for these countries. Finally, for the remaining 16 countries for which neither measure is available, we use data for per capita incomes at PPP in 1997 from the 1998 CIA Factbook (CIA (1998)). We obtain data on infant mortality and adult literacy primarily from the World Development Indicators, supplemented for a few countries from the CIA Factbook. Our data on linguistic variables are based on the data reported by Hall and Jones for their sample of 151 countries, and are extended to our larger sample of 178 countries using Grimes (1996).

Results

Our empirical results show a strong positive causal relationship from improved governance to better development outcomes. In Table 2, we present the results for per capita incomes. The rows of this table report our results for each of the six governance aggregates when entered separately in Equation (4). In the first two columns we report the 2SLS estimates of β and the associated standard errors, for each of these indicators. The most striking feature of these results is that the magnitudes of the estimated coefficients are very large. To see this, recall that our choice of units for governance implies that the standard deviation of governance across countries is equal to one. Therefore, the coefficient on governance can be interpreted as the $100 \times (e^\beta - 1)$ -percent increase in per capita incomes due to a one-standard deviation improvement in governance. The estimated coefficients in the top panel of Table 2 (corresponding to the largest possible sample of countries for each indicator) indicate that a one standard deviation improvement in governance leads to between a 2.5-fold (in the case of voice and accountability) and a 4-fold (in the case of political instability and violence) increase in per capita income. These results clearly indicate that there is a large payoff in terms

¹⁵ The slight difference in timing between the governance and per capita income measures is driven entirely by data availability. Given that the time series variation in these indicators is very small relative to the cross-sectional variation, this discrepancy is unlikely to influence our results.

of per capita income to improvements in governance. In other words, governance does matter.

The next two columns report the results of two key specification tests which are crucial to the plausibility of this simple empirical specification. In the first of these we report the p-value associated with the null hypothesis that the instruments affect income only through their effects on governance. For five out of the six governance measures, we do not reject this null hypothesis, which we interpret as evidence in favour of our identifying assumptions $E[x_j \cdot e_j] = 0$ and $E[x_j \cdot u_j] = 0$.¹⁶

In the next column we summarize the strength of our instruments by reporting the F-statistic from the first-stage regressions of each governance indicator on the instruments. In all cases the F-statistic is highly significant, suggesting that the instruments have significant explanatory power for governance. Although this consideration is irrelevant asymptotically, in finite samples weak instruments may cause the 2SLS estimates to be biased towards the probability limits of the OLS estimates. We therefore take comfort in the finding that these linguistic variables appear to have very good explanatory power for governance. In summary, our specification tests in the full sample of countries suggest that in all cases except graft, we can have considerable confidence that the 2SLS estimator is producing consistent estimates of β and captures the causal effect from governance to per capita income.

In the remaining panels of Table 2 we report the same results for two subsamples of interest. In order to ensure the comparability of the estimated coefficients across indicators, we first restrict the sample to the set of 154 countries which appear in all six governance aggregates. Comparing the estimated coefficients in the first and second panels, we find that the results are quite similar. This reassures us that the differences in estimated coefficients across governance indicators do not simply reflect differences in the samples of countries they cover. In the bottom panel of Table 2 we investigate whether our results are driven only by the differences between developed

¹⁶ Only in the case of graft do we reject this hypothesis, suggesting that our instruments are inadequate in this case. This is not to say that graft is unimportant for economic outcomes. Rather, in this set of countries, we have found it is difficult to find exogenous variations in the causes of graft which make it possible to identify the effects of graft on per capita incomes. We do note that the tests of overidentifying restrictions do pass in the non-OECD sample, and also for the other development indicators as discussed below.

and developing countries. To do this, we re-estimate the six equations excluding all OECD economies from the sample. The results are again qualitatively similar in that we find large and highly significant positive effects of governance on per capita incomes.¹⁷

In Table 3, we report a similar set of results, replacing per capita incomes in turn with the logarithm of infant mortality per thousand live births, and the adult literacy rate. For each of these development outcomes, we re-estimate our regressions for the six governance indicators, in the full set of countries and in the non-OECD sample. Our results are very similar to those for per capita incomes. Improved governance has a strong negative impact on infant mortality, of proportionally the same magnitude as for per capita incomes. Improved governance also leads to significant increases in adult literacy, with a one-standard-deviation increase in governance leading to between a 15 and 25 percentage point improvement in literacy in the full sample of countries. The results of the specification tests are also qualitatively similar.

¹⁷ The only difference is that the F-statistics from the first-stage regressions are somewhat weaker in the non-OECD sample. This raises the possibility that the 2SLS results are biased in the direction of the OLS estimates. We do find, however, that in the full sample of countries the OLS estimates are substantially smaller than the 2SLS estimates. This indicates that the downward biases due to measurement error dominate the possible positive biases from reverse causation or omitted variables. Based on this, we speculate that the 2SLS estimates in the non-OECD sample if anything understate the true effects of governance in this sample.

5. Conclusions

In this paper, we have documented a new database of governance indicators compiled from a variety of existing sources. These indicators report the subjective perceptions of respondents regarding the quality of various aspects of governance. We summarize a subset of these indicators using six aggregate governance indicators capturing six core concepts of governance. Governance, as measured by these indicators, matters a great deal for economic outcomes. We have provided new evidence of a strong positive causal relationship from governance to better development outcomes.

References

- Ades, Alberto and Rafael di Tella (1996). "The Causes and Consequences of Corruption: A Review of Recent Empirical Contributions," *IDS Bulletin*, 27(2).
- Chong, Alberto and César Calderón (1999). "On the Causality and Feedback Between Institutional Measures and Economic Growth". Manuscript, The World Bank.
- Grimes, Barbara (1996). *Ethnologue: Languages of the World*. Summer Institute of Linguistics.
- 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.
- Huther, Jeff, and Anwar Shah (1998). "Applying a Simple Measure of Good Governance to the Debate on Fiscal Decentralization" World Bank Operations Evaluation Department Policy Research Working Paper No. 1894.
- Johnson, Simon, Daniel Kaufmann, and Pablo Zoido-Lobaton (1998). "Regulatory Discretion and the Unofficial Economy". *American Economic Review*. 88(2): 387-392.
- La Porta, Rafael, Florencio López-de-Silanes, Andrei Shleifer, and Robert Vishny (1999). "The Quality of Government". *Journal of Law, Economics, and Organization*. 15(1):222-279.
- Loayza, Norman V. (1996). "The Economics of the Informal Sector: A Simple Model and Some Empirical Evidence from Latin America." *Carnegie-Rochester Conference Series on Public Policy*. 45:129-162.
- Kaufmann, Daniel, and Shang-Jin Wei (1999). "Does "Grease Money" Speed Up the Wheels of Commerce?". NBER Working Paper No. 7093.
- Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobaton (1999). "Aggregating Governance Indicators". Manuscript, The World Bank.
- Kaufmann, Daniel and Pablo Zoido-Lobaton (1999). "Corruption ,Unpredictability and Performance". Manuscript, The World Bank.
- Kaufmann, Daniel, Gil Mehrez, and Sergio Schmukler (1999). "Was the East Asia Crisis Predictable?". Manuscript, The World Bank
- Knack, Stephen Keefer, Philip (1997). "Why Don't Poor Countries Catch Up? A Cross-National Test of an Institutional Explanation". *Economic Inquiry*. 35:590-602.
- Mauro, Paolo (1995). "Corruption and Growth." *Quarterly Journal of Economics*, 110(3): 681-712.

Rodrik, Dani (1997). "TFPG Controversies, Institutions, and Economic Performance in East Asia". NBER Working Paper No. W5914.

Sachs, Jeffrey and Andrew Warner (1995). "Economic Reform and the Process of Global Integration". Brookings Papers on Economic Activity. 1:1-95.

Tanzi, Vito, and Hamid Davoodi (1997). "Corruption, Public Investment, and Growth". IMF Working Paper WP/97/139.

Wei, Shang-Jin. (1997). "How Taxing is Corruption on International Investors?". NBER Working Paper Number 6030.

Table 1: Sources of Quantitative Measures of Governance

Code	Source	Publication	Nature	Country Coverage	Coverage Index
BERI	Business Environment Risk Intelligence	Business Risk Service	Poll	50 mostly developed countries	0.44
CEER	Wall Street Journal	Central European Economic Review	Survey	27 transition economies	0.85
DRI	Standard and Poor's DRI/McGraw-Hill	Country Risk Review	Poll	106 developed and developing countries	0.23 *
EBRD	European Bank for Reconstruction and Development	Transition Report	Poll	26 transition economies	0.85
EIU	Economist Intelligence Unit	Country Risk Service, & Country Forecast	Poll	114 developed and developing countries	0.19 *
FH	Freedom House	Freedom in the World	Poll	172 developed and developing countries	0.03 *
		Nations in Transit	Poll	28 transition economies	0.82
GALLUP	Gallup International	50th Anniversary Survey	Survey	44 mostly developed countries	0.50
GCS	World Economic Forum	Global Competitiveness Survey	Survey	54 developed and developing countries	0.42
GCSA	World Economic Forum	Global Competitiveness Survey, Africa	Survey	23 African countries	0.73
HFWSJ	Heritage Foundation/Wall Street Journal	Economic Freedom Index	Poll	154 developed and developing countries	0.06 *
PERC	Political Economic Risk Consultancy	Asia Intelligence	Survey	11 Asian countries	0.83
PRS	Political Risk Services	International Country Risk Guide	Poll	140 developed and developing countries	0.10 *
WCY	Institute Management Development	World Competitiveness Yearbook	Survey	46 primarily developed countries	0.59
WDR	World Bank	World Development Report	Survey	74 developed and developing countries	0.25 *

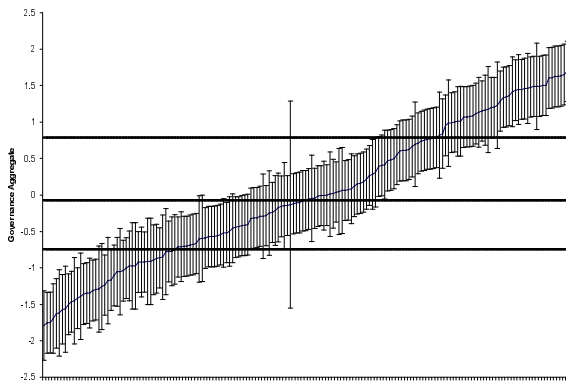
Notes: Poll stands for "polls of experts" and Survey for "survey of entrepreneurs" (except gallup, which is a survey of citizens)

(*) Representative source

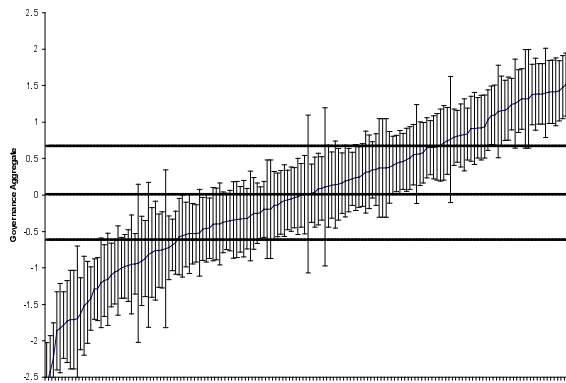
Figure 1: Aggregate Governance Indicators

(i) Process by Which Those in Authority are Selected, Monitored and Replaced

(a) Voice and Accountability

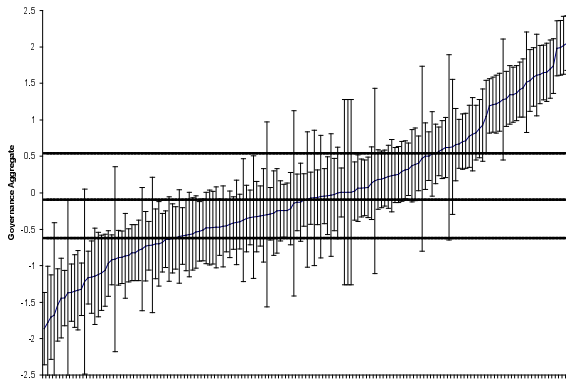


(b) Political Instability and Violence

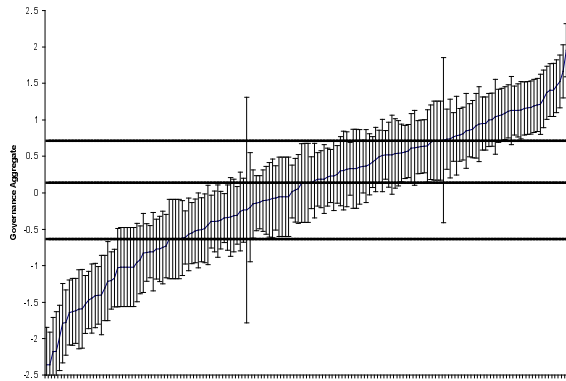


(ii) The Capacity of Government to Manage Resources and Implement Sound Policies

(a) Government Effectiveness

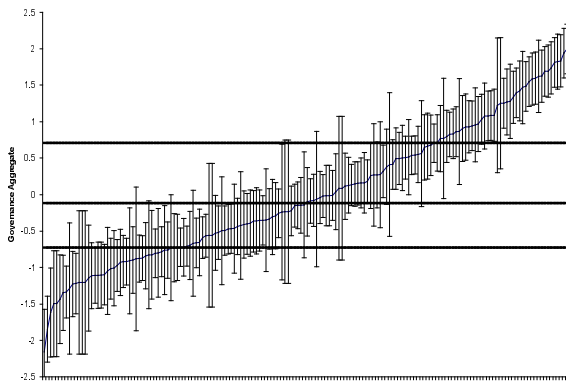


(b) Regulatory Burden

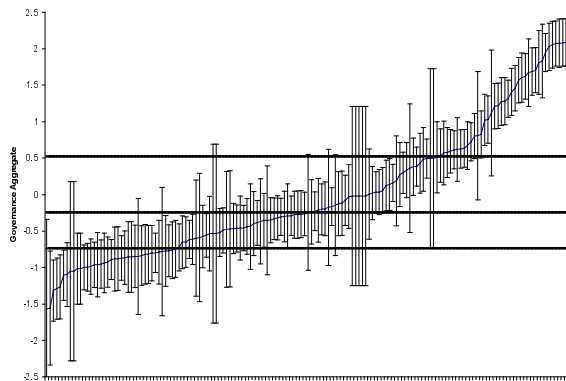


(iii) Respect of Citizens and the State for the Institutions of Society

(a) Rule of Law

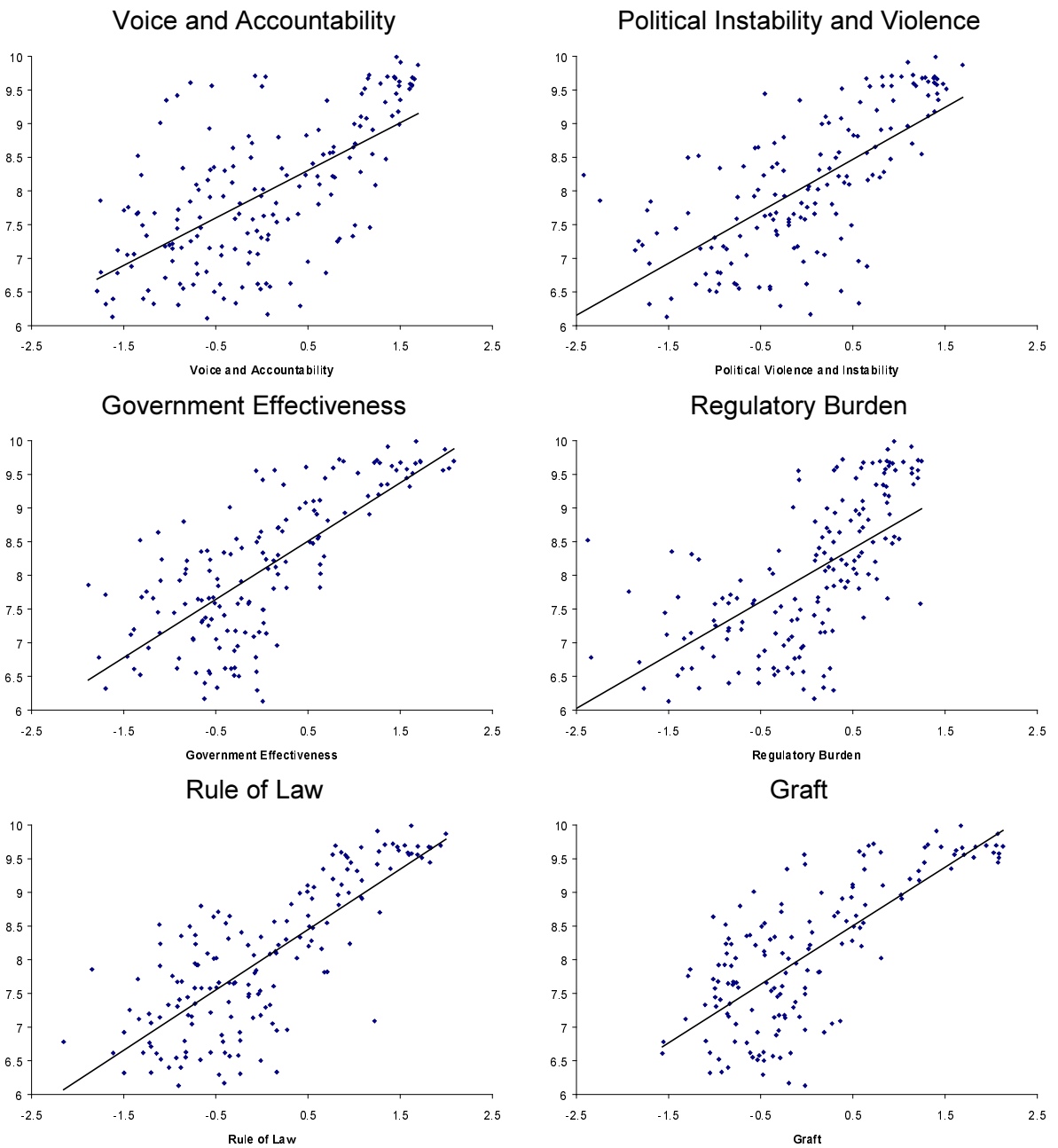


(b) Graft



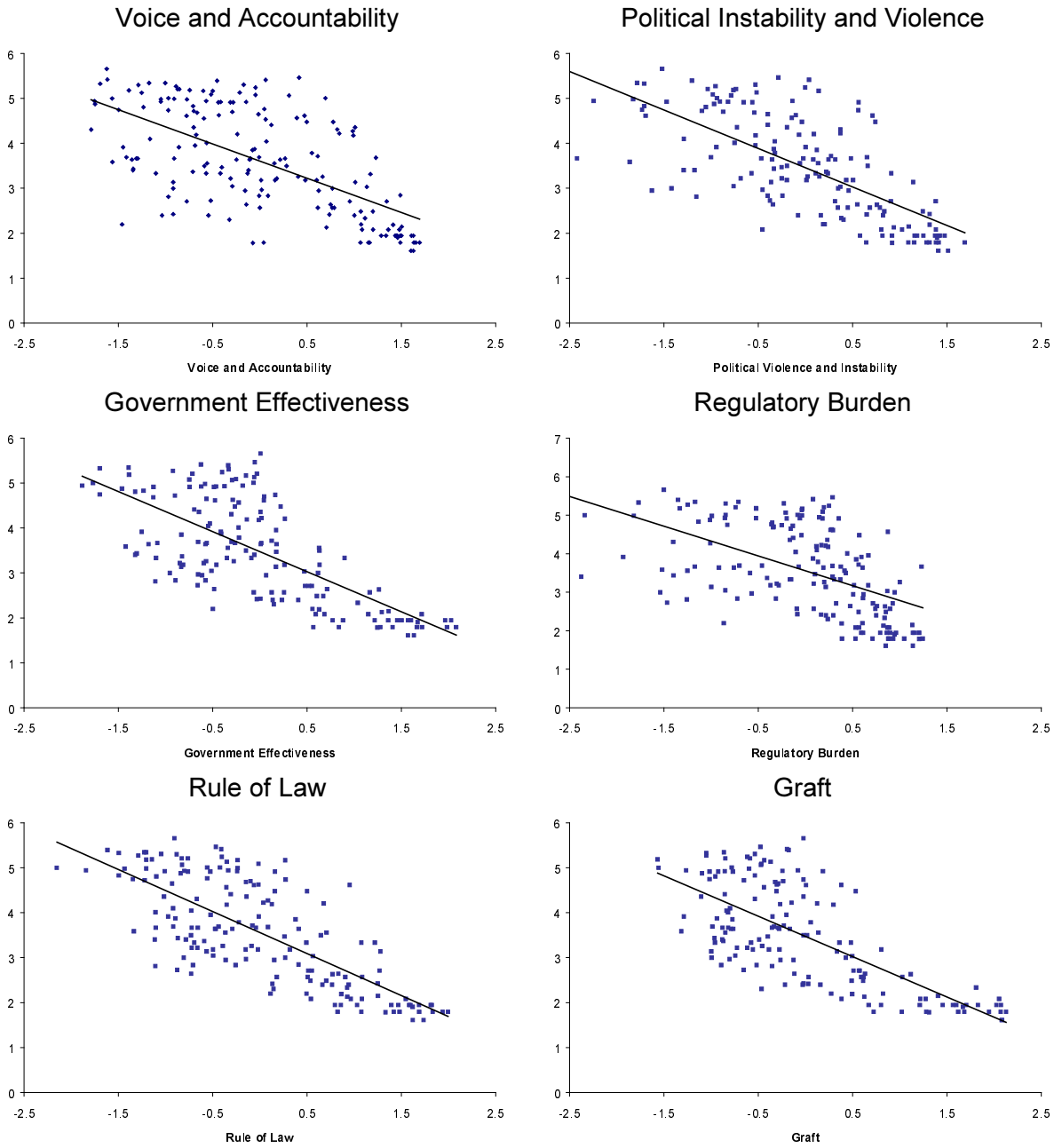
Notes: Each graph reports the value of the governance index (on the vertical axis) for each country (on the horizontal axis), together with a 90% confidence interval. The horizontal lines indicate the first, second and third quartiles of the distribution of governance. Countries are ordered in ascending order from left to right. Higher values of each index correspond to better outcomes.

Figure 2: Governance and Per Capita Incomes



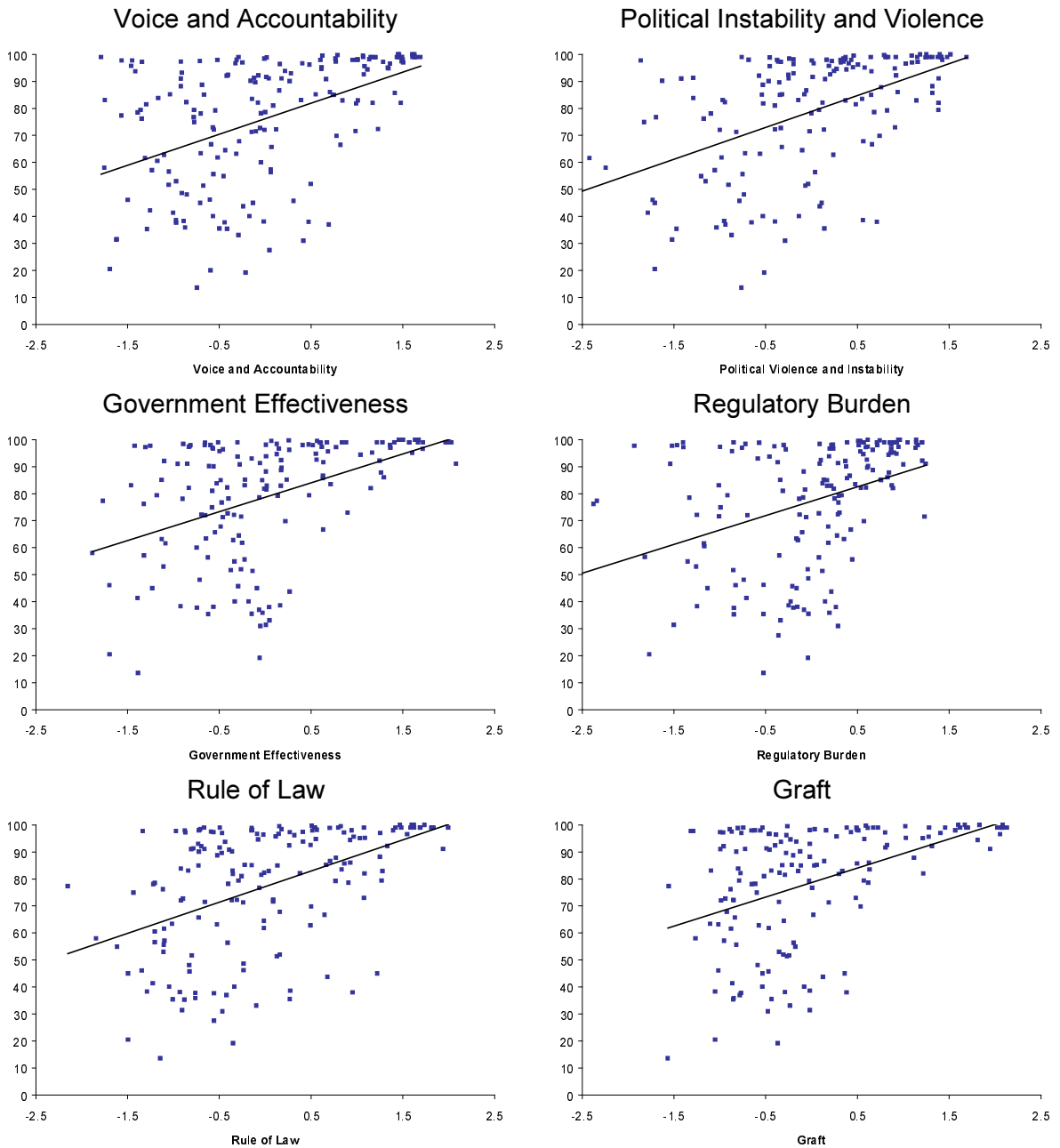
Notes: Each graph plots the indicated governance aggregate (on the horizontal axis) against the logarithm of per capita GDP at PPP (on the vertical axis) for the sample of countries covered by the governance aggregate.

Figure 3: Governance and Infant Mortality



Notes: Each graph plots the indicated governance aggregate (on the horizontal axis) against the logarithm of infant mortality per thousand live births (on the vertical axis) for the sample of countries covered by the governance aggregate.

Figure 4: Governance and Adult Literacy



Notes: Each graph plots the indicated governance aggregate (on the horizontal axis) against the adult literacy rate in percent (on the vertical axis) for the sample of countries covered by the governance aggregate.

Table 2: Regression Results for Per Capita Income

Dependent variable is the logarithm of per capita GDP at PPP

	2SLS		P-Value for	F-statistic	Number of
	β	s.e.	Test of Overidentifying	from first-stage	Observations
			Restrictions	Regression	
Full Sample					
Voice and Accountability	0.888	0.109	0.167	23.025	173
Political Instability and Violence	1.394	0.177	0.485	6.301	155
Government Effectiveness	1.102	0.128	0.127	12.080	156
Regulatory Burden	1.133	0.143	0.874	18.413	166
Rule of Law	1.335	0.179	0.122	7.983	166
Graft	0.990	0.112	0.018	12.069	155
Sample of Countries Appearing in All Six Aggregates					
Voice and Accountability	0.907	0.123	0.175	17.559	154
Political Instability and Violence	1.390	0.177	0.491	5.979	154
Government Effectiveness	1.095	0.130	0.106	11.177	154
Regulatory Burden	1.085	0.155	0.792	16.998	154
Rule of Law	1.237	0.171	0.071	6.952	154
Graft	0.977	0.111	0.017	11.567	154
Non-OECD Sample					
Voice and Accountability	0.685	0.143	0.093	14.259	149
Political Instability and Violence	1.811	0.532	0.712	1.190	131
Government Effectiveness	1.561	0.452	0.486	2.552	132
Regulatory Burden	0.880	0.169	0.990	11.250	142
Rule of Law	2.079	0.670	0.365	1.227	142
Graft	1.192	0.317	0.096	2.227	131

Table 3: Regression Results for Other Development Outcomes

	2SLS β	s.e.	P-Value for Test of Overidentifying Restrictions	F-statistic from first-stage Regression	Number of Observations
Dependent Variable is Infant Mortality Per 1000 Live Births					
Full Sample					
Voice and Accountability	-0.847	0.118	0.324	23.025	173
Political Instability and Violence	-1.408	0.212	0.610	6.301	155
Government Effectiveness	-1.124	0.165	0.283	12.080	156
Regulatory Burden	-1.101	0.142	0.816	18.413	166
Rule of Law	-1.361	0.218	0.126	7.983	166
Graft	-1.025	0.140	0.013	12.069	155
Non-OECD Sample					
Voice and Accountability	-0.636	0.169	0.309	14.259	149
Political Instability and Violence	-1.842	0.722	0.966	1.190	131
Government Effectiveness	-1.573	0.590	0.853	2.552	132
Regulatory Burden	-0.812	0.186	0.790	11.250	142
Rule of Law	-2.211	0.977	0.445	1.227	142
Graft	-1.495	0.533	0.149	2.227	131
Dependent Variable is Adult Literacy Rate in Percent					
Full Sample					
Voice and Accountability	15.102	1.690	0.164	23.025	173
Political Instability and Violence	26.140	5.242	0.611	6.301	155
Government Effectiveness	20.445	3.911	0.336	12.080	156
Regulatory Burden	20.902	2.809	0.932	18.413	166
Rule of Law	25.167	5.354	0.220	7.983	166
Graft	18.602	3.544	0.080	12.069	155
Non-OECD Sample					
Voice and Accountability	15.857	2.688	0.103	14.259	149
Political Instability and Violence	55.701	20.129	0.791	1.190	131
Government Effectiveness	44.342	13.794	0.987	2.552	132
Regulatory Burden	22.281	4.224	0.672	11.250	142
Rule of Law	54.679	21.820	0.473	1.227	142
Graft	36.789	8.252	0.269	2.227	131

End of main text

This paper has been split into main text and appendix segments due to downloading constraints.

Appendices for this paper can be found at
http://www.worldbank.org/wbi/governance/gov_pdfs/govmatrs-annex.pdf.