

Governance Indicators, Aid Allocation, and the Millennium Challenge Account

Daniel Kaufmann and Aart Kraay

The World Bank
December 2002

I. Introduction

There is widespread consensus that development assistance works best when it is targeted towards countries with relatively sound and/or improving policies and institutions. Recognizing this, bilateral and multilateral donors are increasingly trying to direct their assistance towards recipients with these characteristics.¹ Most recently and prominently, on November 25, 2002 the U.S. government released details of how it plans to allocate funds from the new \$5 billion per annum Millennium Challenge Account (MCA) towards countries that “govern justly”, “invest in people”, and “promote economic freedom”.² This represents a major policy shift by a donor in moving to an allocation criteria which places governance issues center stage, and which relies on a highly transparent and objective empirical allocation criteria.

The proposed criteria for country eligibility draw heavily on a number of cross-country measures of the quality of governance, including several that we have constructed as part of an ongoing project in the research department of the World Bank and the World Bank Institute.³ We note at the outset that these research indicators, as well as the views expressed here, do not necessarily represent the official views of the World Bank. In this brief note, we describe these governance indicators and the MCA’s proposed use of them. We then offer some thoughts and suggestions on what we regard as unresolved issues regarding the use of this kind of data as a tool to achieve the unquestionably important objective of more effective aid allocation.

II. Background on Governance Indicators and MCA Allocation

What are aggregate governance indicators? Based on data provided by a broad range of stakeholders around the world, we have constructed six cross-country

¹ For example, for several years the World Bank has used its own internal assessments in allocating resources from the Bank’s concessional lending facility, the International Development Association (IDA).

² This document can be found at <http://www.worldbank.org/wbi/governance/mca.htm>, where feedback is also welcome. See also http://www.cgdev.org/nv/features_MCA.html for several papers by scholars at the Center for Global Development discussing a variety of complementary issues in the design of the MCA.

³ The detailed dataset for all indicators and countries, as well as links to the background research papers, is available at: <http://www.worldbank.org/wbi/governance/govdata2001.htm>.

indicators of measuring six dimensions of the quality of governance: *voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption*. We have constructed these indicators for two periods (1997/98, and 2000/01) and we will shortly be releasing an update for 2002. Depending on the particular component of governance we measure, the 2000/01 indicators cover between 157 and 173 countries, and in the forthcoming update we plan to cover nearly 200 countries for the 2002 indicator (to be available in early 2003).

In order to improve reliability and country coverage, each of these six governance research indicators combines a large number of underlying measures of perceptions of governance. In the 2000/01 indicator, we drew on 194 separate measures compiled by 17 different sources obtained from a variety of international organizations, survey institutes, risk-rating agencies, and think-tanks. Table 1 gives a list of the sources we have used in this exercise. The statistical methodology we utilize in arriving at these composite governance research indicators not only gives us estimates of governance but also margins of error for each country. Although we stress that the aggregate indicators are more reliable in a statistical sense than any individual source, these margins of error can in many cases still be quite substantial. We discuss the implications of this in more detail below.

Table 1: Sources of Governance Data

| <u>Source</u> | <u>Publication</u> |
|--|--|
| Business Environment Risk Intelligence | Business Risk Service |
| Columbia University | State Capacity Project |
| Economist Intelligence Unit | Country Risk Service |
| European Bank for Reconstruction and Redevelopment | Transition Report |
| Freedom House | Nations in Transition |
| Freedom House | Freedom in the World |
| Gallup International | Gallup Millennium Survey |
| Heritage Foundation/Wallstreet Journal | Economic Freedom Index |
| Institute for Management and Development | World Competitiveness Yearbook |
| Latinobarometro | Latinobarometro Surveys |
| Political Economic Risk Consultancy | Asia Intelligence |
| Political Risk Services | International Country Risk Guide |
| PriceWaterhouseCoopers | Opacity Index |
| Standard and Poor's DRI McGraw-Hill | Country Risk Review |
| World Bank | Business Enterprise Environment Survey |
| World Bank | World Business Environment Survey |
| World Economic Forum | Global Competitiveness Report |

How will governance indicators be used to determine MCA eligibility? The proposed MCA allocation rule is designed to ensure that MCA funds will be allocated to low-income countries with relatively sound policies and institutions. A group of 74 countries that are eligible for concessional IDA lending from the World Bank, and which have per capita incomes less than \$1435 in 2001, will potentially be eligible for MCA

funds in its first year.⁴ According to the MCA eligibility rules, this set of countries will be rated according to 16 performance criteria covering three dimensions of performance: “governing justly” (6 criteria), “investing in people” (4 criteria), and “promoting economic freedom” (6 criteria). These criteria are listed in Table 2. Four of the governance research indicators we have constructed (voice and accountability, government effectiveness, rule of law, and control of corruption) have been proposed as performance indicators under the MCA’s “governing justly” performance dimension, with the remaining two for this dimension being measures of civil liberties and political rights constructed by Freedom House. In addition, a fifth governance research indicator, regulatory quality, is included under “promoting economic freedom”.

Table 2: MCA Performance Criteria

Governing Justly:

- Civil Liberties (Freedom House)
- Political Rights (Freedom House)
- Voice and Accountability (World Bank)
- Government Effectiveness (World Bank)
- Rule of Law (World Bank)
- Control of Corruption (World Bank)

Investing in People:

- Public Primary Education Spending as Percent of GDP (World Bank/national sources)
- Primary Education Completion Rate (World Bank/national sources)
- Public Expenditures on Health as Percent of GDP (World Bank/national sources)
- Immunization Rates: DPT and Measles (World Bank/UN/national sources)

Promoting Economic Freedom:

- Country Credit Rating (Institutional Investor Magazine)
- Inflation (IMF)
- 3-Year Budget Deficit (IMF/national sources)
- Trade Policy (Heritage Foundation)
- Regulatory Quality (World Bank)
- Days to Start a Business (World Bank)

In order to qualify for MCA assistance, countries must (a) be in the top half of all potentially eligible countries according to the *control of corruption* rating from the governance research indicators, and (b) must be in the top half of all potentially eligible countries on at least half of each of the performance criteria under each of the three dimensions of performance. This rule is designed to ensure that resources are

⁴ A number of countries with per capita income greater than \$1435 are currently eligible for IDA under the small island economies exception, but presumably these will not initially be eligible for the MCA. The group of 74 countries is based on data on IDA eligibility available at www.worldbank.org/ida, and per capita gross national income in US dollars in 2001 using Atlas exchange rates available in the World Bank’s World Development indicators. There will likely be differences between this list and the official list compiled by the MCA administration due to differences in data sources and updates, etc. In later years, eligibility will expand to all countries with per capita incomes less than \$1435, and later to all countries with per capita incomes less than \$2975.

channeled towards countries that are performing well in a variety of dimensions of governance, and in which corruption especially is relatively low. Given the abundant evidence of the importance of good institutions and policies for growth, development, and aid effectiveness, this type of allocation rule is certainly warranted. Moreover, an objective and monitorable set of criteria for determining MCA eligibility is highly desirable, both in terms of the process of aid allocation, and also in terms of creating clear incentives among potential recipients of this aid.

In the remainder of this note, we highlight six points for consideration in the use of this type of data as a basis to allocate aid. It is timely to have an open discussion on key unresolved issues at this juncture, given that some important details about how MCA allocation rules will work are still to be finalized, and that another round of data collection is still to take place (therefore potentially allowing for adaptations and improvements in the next round of performance indicators measurement).

We also note that a debate on these particular challenges in applying the clear rules set out by the MCA has much broader implications for the wider donor community, as various donors are considering similar paths in allocating aid -- and also because many of the points we bring up on the challenges of moving from measured indicators into actual allocations apply much more broadly as well. In this sense, the focus on MCA rules in this paper ought to be seen as illustrative of a broader set of problems and challenges. Further, it is important to keep in mind that our focus here on *governance indicators* derives from the fact that this has been our research area -- yet we are aware that many of the points made below do apply for other indicators (outside of core governance) as well.

III. Six Issues in the Use of Governance Indicators to Allocate Aid

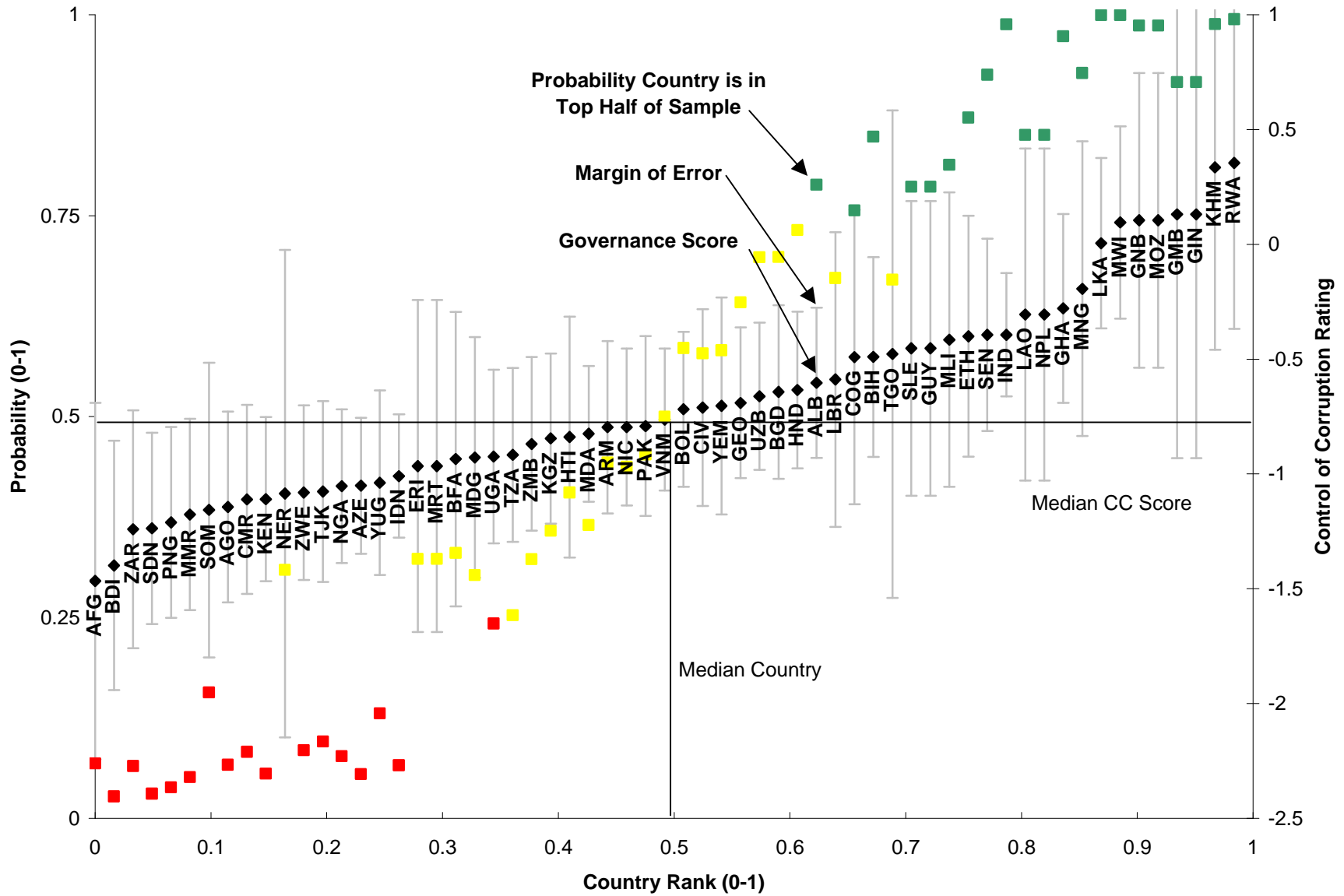
1. Governance indicators have substantial margins of error, so that special empirical scrutiny is needed for borderline cases.

The key point here is that these substantial margins of error mean that for many countries it is difficult to assign them with a high degree of confidence to a definitive performance category according to their estimated level of governance.⁵ This point applies to any of the MCA criteria, but for concreteness consider the control of corruption indicator where, unlike many of the MCA criteria, we have explicit margins of error. While we certainly support the approach that corruption should be an important factor in allocating aid, it is important to emphasize that a simple “in-or-out” rule runs the risk of misclassifying some countries precisely because margins of error are not trivial. This possibility is recognized in the MCA fact sheet, which notes that Board of the MCA will be “...encouraged to identify for special transition support a small number of countries that barely miss the list of better performers.”

Figure 1 emphasizes the importance of margins of error. We rank the 61 out of 74 potential MCA countries for which we also have data on corruption in our 2000/01 indicator. For each country, we show the estimated corruption rating as a black diamond, and we show the margins of error for each country as a vertical line. The interpretation of these margins of error is the following: we can be 90% confident that

⁵ This point is echoed by Steve Radelet in his comments on the MCA allocation rule (http://www.cgdev.org/nv/features_MCA.html).

Figure 1 – Margins of Error and Governance Rankings



governance is in the range indicated by the thin vertical line for each country. The first point to note from Figure 1 is that there are few countries for which the (admittedly stringent) 90% confidence range is entirely in either the bottom or top half of the sample. At the bottom end, only Burundi and Sudan have confidence ranges entirely in the bottom half of scores, and at the top end, 8 countries including India, Malawi and Cambodia are fully in the top half of scores.

For the majority of countries there is a non-trivial probability that they could be mistakenly classified in the bottom half of the sample -- when a 'perfectly accurate measure' (which does not exist in reality) would have indicated that they should be in the top half, and vice versa. To illustrate this more precisely, for each country our methodology allows us to calculate the probability that a country's true unobserved level of governance falls in the top half of the sample. These probabilities are indicated as squares in Figure 1. Not surprisingly for the worst-rated countries, the probability they could fall in the top half of the sample is close to zero. Similarly, the best rated countries almost certainly belong in the top half. However, there is a large intermediate range of countries where there is a non-trivial probability that they belong in either the top or bottom half of the sample, for example ranging from around 0.25 to 0.75. Borrowing colors from a traffic light, we have color-coded the first group as red (less than 25% chance that they are mistakenly classified in the bottom half), the second as green (more than 75% chance that they actually belong in the top half), and the intermediate group as yellow.

This "traffic light" approach highlights the challenge of assigning countries to performance categories -- particularly the substantial number that fall in the "yellow light" category. While the 21 "green light countries" most likely belong in the top half, and the 17 "red light" countries in the bottom half, the 23 intermediate "yellow light" countries are a more difficult case. These countries are difficult to assign to categories simply because the available cross-country data is not sufficiently informative and/or there is disagreement between the underlying sources. The importance of addressing the difficulty of assigning the "yellow light" countries to either the "green" or "red" categories points to the importance of relying on additional sources of information on which to base MCA eligibility decisions for this group. Below we discuss one potentially valuable source of such information -- in-depth country-specific governance and anti-corruption diagnostics such as those that the Bank has been implementing in about thirty countries worldwide already. This also underscores the need for a certain degree of flexibility in the MCA allocation rule, and importantly, that this flexibility should be symmetric. Not only should countries that "barely miss the list of better performers" be given special consideration as currently proposed in the MCA fact sheet, but in addition countries that barely make the list of better performers should also merit further scrutiny.

2. Margins of error can be reduced by relying on more sources of information.

In light of our previous point, measures with smaller margins of error are desirable because they reduce the risk of misclassifying countries. The margins of error in these aggregate governance indicators reflect two factors -- the number of sources of information available for each country, and the quality of the underlying source itself. We illustrate this in Figure 2, again using data from the control of corruption indicator, to show the relationship between margins of error and the number of sources available for each country. This relationship is strongly negative -- margins of error for countries with four or five sources of data are half as large as those for countries with only one source of information.

Figure 2 – Margins of Error for Individual and Aggregate Corruption Indicators

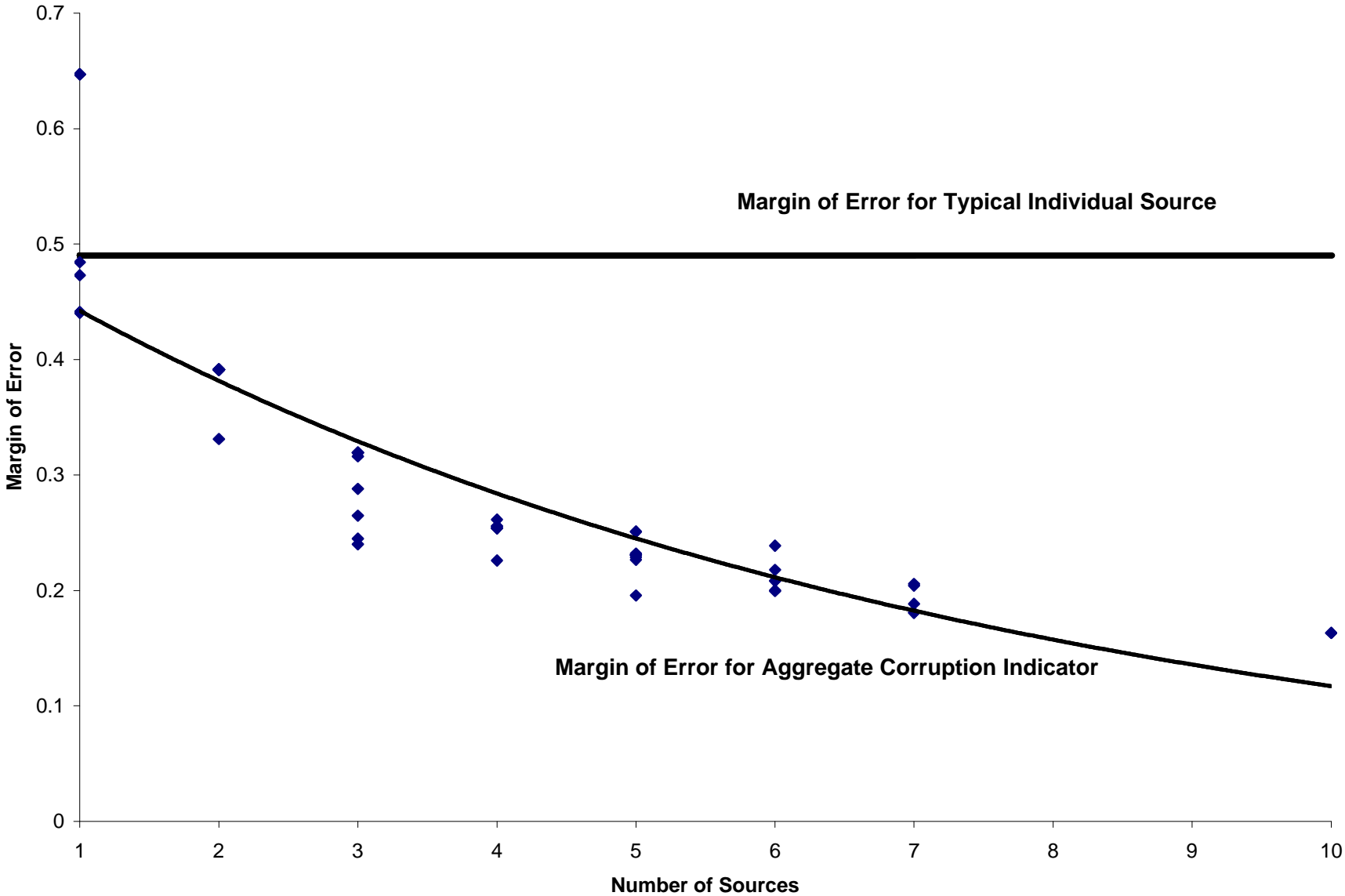
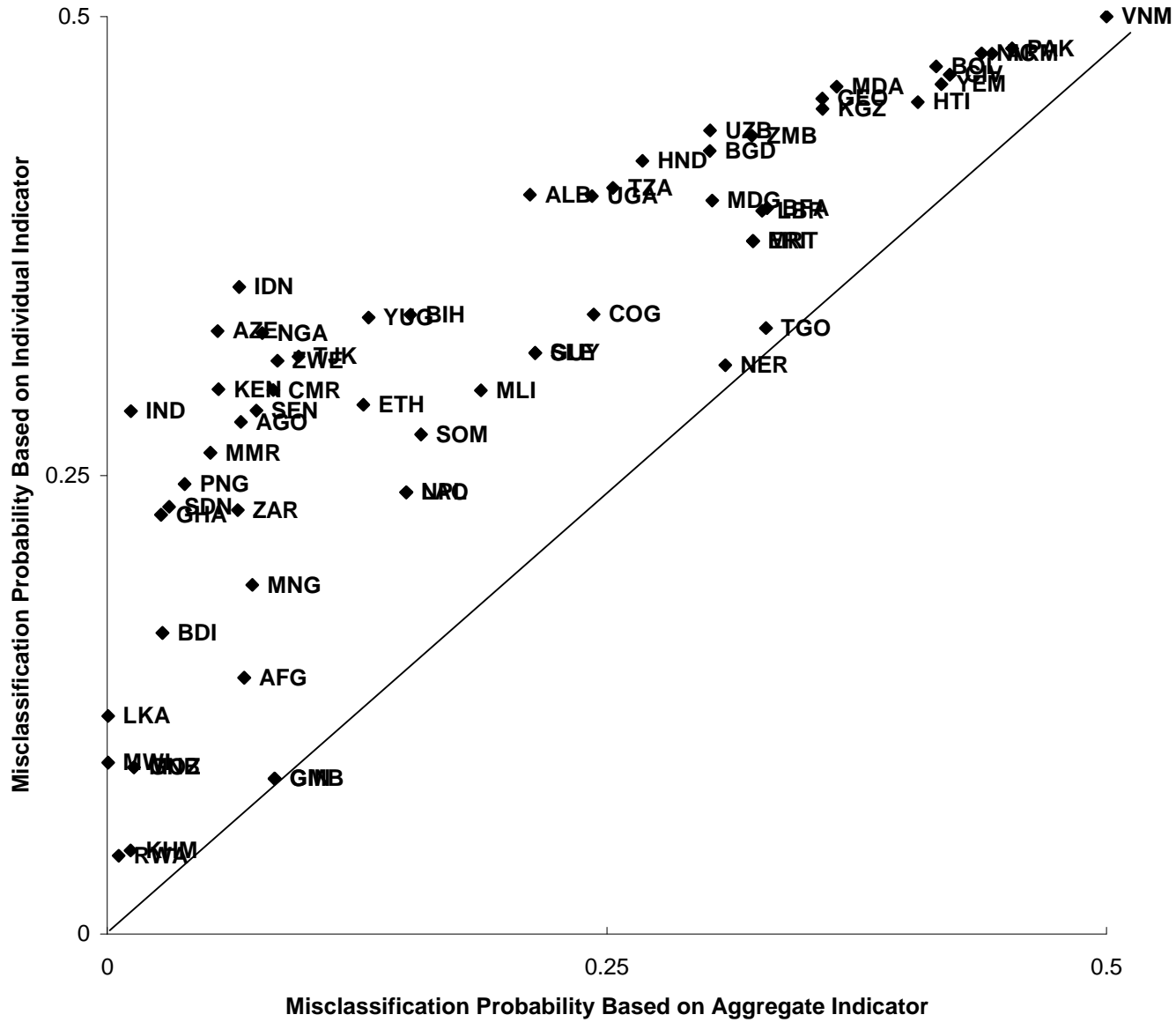


Figure 3 – Classifying Countries Using Aggregate and Individual Indicators



An important consequence of this point is that using individual (as opposed to aggregate) sources of information on governance to classify countries should be done with an even greater abundance of caution. Figure 3 (above) shows how the probability of misclassifying countries increases if we were to rely on only a typical single source of corruption data. On the horizontal axis, we graph the probability that a country classified in the bottom (top) half of the sample is actually in the top (bottom), based on the *aggregate* corruption indicator. On the vertical axis, we plot the same probability, but instead assuming that we were to rely on a typical *single* source of corruption data out of the many that we use. For countries with very good or very bad scores, it makes little difference whether we rely on individual or aggregate indicators. However, for most every intermediate country (unless the aggregate indicator has only relied on one individual source, as is the case in very few cases), the probability of misclassification is much higher if we use individual as opposed to aggregate indicators. In fact, the group of 23 “yellow light” countries identified above using the aggregate indicator nearly doubles to 44 countries if we were to rely only on a single source of corruption data.

3. Margins of error are not unique to subjective indicators.

Although we have emphasized the importance of the explicit margins of error that can be calculated from aggregate governance indicators based on subjective perceptions of governance, it is important to note that margins of error are not unique to subjective data. All of the 16 performance indicators proposed for the MCA (and virtually any indicator) are vulnerable to measurement error of two sorts:

- Indicators may at best be proxies for good performance, even if they can be measured accurately. Consider for example per capita spending on health or education. While this spending can in principle be measured well, it will not be perfectly correlated with the performance criteria of “investing in people” to the extent that these resources are poorly targeted or inefficiently spent. The same is true for “objective” indicators of other dimensions of governance. For example, low reported crime rates, while easily measurable, may be a poor proxy for good rule of law if police are incompetent or mistrusted by the people. In a similar vein, high rates of incarceration may be a reflection of relatively good or poor rule of law. A rigorous conceptual rationale is thus always required in selecting appropriate indicators which do capture the essence of the concepts being measured (such as governance).
- Even when there is a good correspondence between what is being measured and the underlying concept of interest, measurement error remains pervasive in all kinds of official and/or otherwise ‘objective’ socioeconomic data. In most cases, this is unfortunately not explicitly recognized in today’s statistics, but it is clearly present, even in industrialized countries such as the United States. For example, in mid-2002, the U.S. Bureau of Economic Analysis had produced three estimates of GDP growth for the fourth quarter of 2001, ranging from 0.2% to 1.7%.⁶ It is not hard to imagine that measurement errors can be much more substantial in many developing countries with less advanced statistical systems.

⁶ Ronald Wirtz. Federal Reserve Bank of Minneapolis, “GDP: Understanding News from Noise”. *The Region*, June 2002. (<http://minneapolisfed.org/pubs/region/02-06/gdp.cfm>). It is noteworthy that nowadays rigorous articles on margins of error about official statistics as this one are the

Since measurement error and margins of error are pervasive, and are not unique to subjective governance data, it is important that allocation rules take these margins of error as seriously as possible. One attractive feature of the proposed MCA allocation rule is that it requires good performance across a range indicators, which may help to average out errors in individual categories.

4. Country coverage of governance indicators is important.

In our 2000/01 governance indicators we covered 61 of the 74 potential MCA eligible countries, and in our expanded and updated indicators for 2002, we expect to cover virtually all of them. Coverage for many of the other proposed MCA indicators is less comprehensive. This raises a subtle but important potential difficulty. Suppose that the 13 countries for which corruption data is missing tend to be countries with worse-than-average corruption scores. Then several of the countries that rank in the bottom half of the list of 61 countries for which data are available would not rank in the bottom half of all 74 potential MCA countries. These countries would in effect be penalized by the absence of data for the 13 missing countries, let alone potentially misclassifying the countries with missing data themselves. If on the other hand the 13 omitted countries had better than average corruption scores, then some of the countries in the top half of the list of 61 would not find themselves in the top half of the list of all 74 countries.

The potential difficulty this presents is worse for indicators that cover fewer countries. The most extreme example is probably the 3-year average budget deficit requirement proposed for the MCA. Using published data in the IMF's Government Finance Statistics, we were able to identify only 13 out of 74 potential MCA countries with complete budget deficit data for 1998-2000. To the extent that data availability and better performance go hand-in-hand, this raises the possibility that countries may incorrectly be classified in the bottom half of the sample only because many worse-performing countries have no data at all. We should also note in this context that the breadth of coverage does not only have a 'spatial' dimension (the world), but a time dimension as well: *timeliness* of the information gathered for the indicators is also very important in order to lower the margins of error and probability of misclassification. Some of the proposed indicators to be used under the MCA (not in governance) unfortunately appear to exhibit substantial lags, a challenge which will also need to be addressed – alongside the expansion of country coverage.

5. Measuring improvements in governance is important, yet difficult.

There are at least two reasons why measuring *changes over time* in governance is also important in allocation rules.

- It is useful to be able to measure changes in order to shift the allocation of aid resources over time towards countries with greater improvements in governance, and away from those with deterioration.
- The quality of governance in a country reflects a complex array of factors. Important among these are historical influences. Recent research into the origins

exception, contrasting the explicit treatment of these issues fifty years ago by academic giants such as Simon Kuznets (on national accounts of the US) and Van Morgenstern.

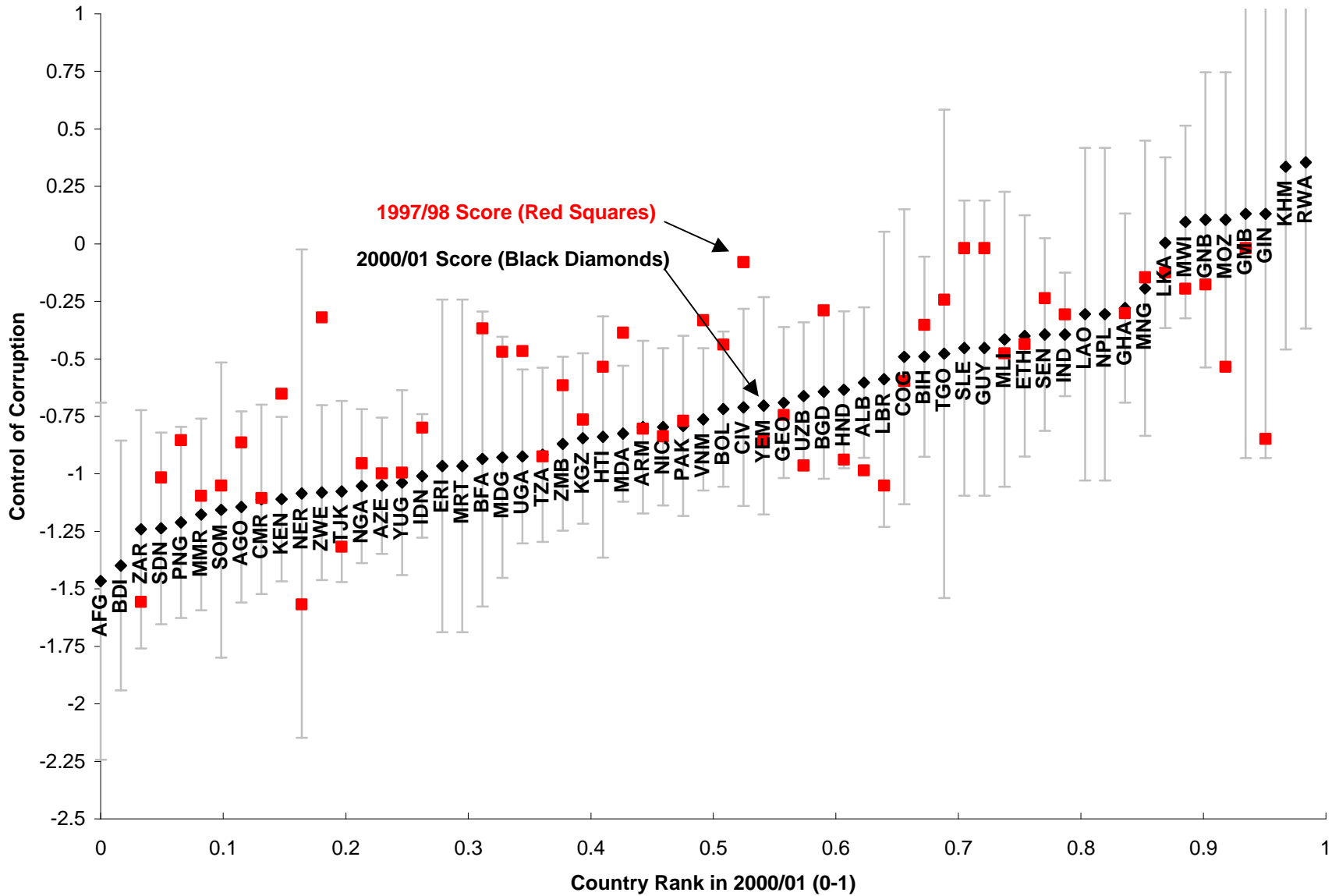
of good and bad governance has documented the importance of colonial origins, geographical variables, and initial factor endowments and the interactions between these in determining current levels of institutional quality. In light of this, there is an argument that countries should not be penalized for factors outside their control in aid allocation rules, and that countries should also be rewarded for recent improvements in governance. Consequently, an allocation rule needs to also consider the trends in governance, rather than relying solely on levels.

Despite its importance, measuring changes in governance over time is very difficult, and especially over relatively short periods of a couple of years. Figure 4 illustrates this using the aggregate corruption indicator, which is available for 1997/98 and 2000/01. As in the previous graph, the black diamonds and vertical lines show the 2000/01 corruption estimates and margins of error for the potential MCA countries. The red squares show the corresponding scores for 1997/98. This figure highlights the fact that changes over time in these cross-country indicators are for most countries generally small relative to margins of error. Only a handful of countries have corruption scores for 1997/98 that lie outside the (90% confidence level) margins of error for 2000/01. These are the estimated worsenings in corruption in Kenya, Zimbabwe, Vietnam, and Cote d'Ivoire.

While we can be reasonably confident that these measures give an indication of the direction of change in governance, this points to the importance of caution in interpreting the significance of these changes, especially over very short periods of only a few years. Many dimensions of governance are likely to change fairly gradually over time. As more data becomes available (for example, the planned update of our governance indicators with retrospective data for 1995 as well) it may be possible to identify more substantial trends in governance over time.

Nevertheless, while this type of cross-country corruption data can identify large improvements or declines, it is likely to remain a blunt instrument for measuring more gradual progress. This points to the importance of using alternative sources of information, not only to improve measures of the *level* of governance as discussed above, but also about *trends* in governance.

Figure 4 – Changes over Time in Governance



6. In-depth country diagnostics can usefully complement cross-country data

We have argued that cross-country indicators can be informative for assessing broad *levels* of governance performance for countries, as long as any false claim of precision in rankings are avoided. We have also seen that these measures can provide some indication of trends in governance. However, in both areas, measurement is sufficiently imprecise that further complementary efforts are required.

In this context, the detailed country diagnostic surveys designed by the World Bank in the past few years can help in this respect. These Governance and Anti-Corruption (GAC) diagnostics rely on in-depth, country-specific surveys of thousands of: i) public service users; ii) firms, and, iii) public officials, in order to gather specific information about institutional vulnerabilities within a country. To date, they have been implemented or are in process in 30 countries. These separate surveys are carried out by local non-governmental institutions, and permit triangulation and consistency checks for the results across respondent categories, while probing in much more detail into a broad array of governance issues within each country, including the time dimension. Selected questions are put to the different respondent stakeholders in terms of the trends over the past two (and at times also five) years for a variety of governance variables. Furthermore, these in-depth baseline diagnostics are intended to provide an initial benchmark, out of which a periodic monitoring effort is expected to ensue. As periodic monitoring of the core issues and questions take place on many different dimensions, further validated by the trend questions to respondents in each survey, a better assessment of trend lines within each country is expected to result.

These in-depth diagnostic and monitoring surveys have a very important complementary benefit to another gap of cross-country indicators: informing policy-makers on priorities for formulating actions and reforms. An innovation in these surveys has been their emphasis on “unbundling” governance and corruption into more detailed and specific dimensions. This has helped to highlight the causes, consequences, and costs of various forms of misgovernance, and has shown how there can be wide variation in institutional quality across institutions within a particular country. This in turn, has helped shaped strategic and reform priorities in many countries where it has been applied.⁷ Furthermore, since these are in effect ‘self’-diagnostics because local organizations implement it by seeking detailed information from local stakeholders, which is then subject to a participatory coalition-building and collective action process with all key stakeholders in society, these GAC diagnostics are a tool synthesizing ‘local knowledge’ instrumental in promoting transparency, participation, domestic ownership of the challenges and action programs

Consequently, given the virtues and characteristics of these in-depth GAC diagnostic surveys in complementing the pitfalls of broad cross-country indicators (particularly in providing information for countries whose estimated levels of governance place them in a ‘yellow light’ category, for assessing short-run trends, and in providing an input to strategy formulation and informing country policy reformers), a recommendation would be to have a country carry out this GAC diagnostic and monitoring methodology on a periodic basis. This could be done, say, every two years

⁷ For details on governance diagnostic instruments and country reports, visit <http://www.worldbank.org/wbi/governance/tools.htm>.

or so (and not always as extensively as in the first baseline diagnostic, the aim being to provide for a more flexible and rapid monitoring tool which is institutionalized).

The recommendation would be that in the first instance the in-depth country diagnostic focus is given to countries in the 'yellow light' category discussed above, given their higher likelihood of misclassification, and also due to the benefits that such self-diagnostics may provide in internally helping formulate and promoting good governance reforms – thereby potentially helping a country move to the 'green light' category, and farther out of danger of falling to the 'red light' zone. In fact, any country intent in carrying out its own self-diagnostic utilizing this type of rigorous framework, as well as embracing transparency and participation in its process, could already receive additional consideration in the allocative classification decisions, since the very fact and process of embarking in these GAC diagnostics (irrespective of what the data collected may reveal), is in itself an important pro-governance and pro-transparency move.

IV. Conclusions

Aid works more effectively in a better – and improving-- policy and institutional environment. Improving the allocation of aid to recognize this fact can do much to improve the effectiveness of development assistance. Thus, efforts to utilize governance indicators for more objective and transparent eligibility criteria ought to be welcomed. At the same time, directing aid towards countries with good governance raises measurement challenges, which need to be addressed in order to enhance effectiveness in the use of indicators for aid allocation eligibility criteria. In this brief note we have highlighted six of these challenges, illustrating our points with the use of governance indicators in the proposed MCA allocation rule, and suggesting some ways forward. One way to complement the admittedly imperfect information provided by worldwide governance indicators is to carry out in-depth, country-focused governance diagnostic surveys for selected countries. By focusing on intermediate countries, this can reduce the risk of misclassification and would also help with internal ownership and action program implementation. This complementary approach to obtaining additional information is particularly important if, as we argue, measurement of the *changes* in governance over time is to be important as well in implementing eligibility criteria (alongside the use of *level* estimates in such criteria).

While our focus in this note has been on governance indicators, we note in conclusion that the set of six point points made, as well as the recommendations, apply more broadly to other indicators and their interpretation as well.

Feedback welcomed on this draft discussion paper at <http://www.worldbank.org/wbi/governance/mca.htm>.