CHAPTER FOURTEEN

Choosing and Using Performance Criteria

Robert Klitgaard, Johannes Fedderke, and Kamil Akramov

Introduction

The Volcker Commission calls for performance-driven public management. Which performance measures should be chosen? And how should the chosen measures be used?

This chapter looks at a current example, the Millennium Challenge Account, but its goal is more general. It shows how to use performance measures to select a few among many candidates (countries, agencies, programs, people) for special benefits. Choosing and using performance measures has four effects:

1. Allocative efficiency
2. Distributional effects
3. Incentive effects
4. Fundraising effects

Those choosing and using performance measures should analyze all four effects—something that is apparently seldom done in practice or in the academic literature.

Governance and Development

This chapter illustrates the use of governance measures to allocate additional foreign aid. In February 2003, President George W. Bush sent Congress a bill that will increase foreign aid by 50 percent over

the next three years by creating a Millennium Challenge Account (MCA) for a select group of poor countries. In March 2002, President Bush said the MCA will

> reward nations that root out corruption, respect human rights, and adhere to the rule of law . . . invest in better health care, better schools and broader immunization . . . [and] have more open markets and sustainable budget policies, nations where people can start and operate a small business without running the gauntlets of bureaucracy and bribery.[1]

By early 2004, the Bush administration had identified 63 countries eligible to compete for MCA funds because their per capita income (GDP p.c.) was below $1,415 and they were not "sponsors of terrorism." These countries were then rated on 16 performance measures.[2] To receive MCA funds, a poor country has to score above the median on the anticorruption indicator and above the median in half the indicators in each of three domains of performance.[3]

---

[1] Remarks by the President on Global Development, March 14, 2002, Washington, DC: Office of the Press Secretary (available at http://usinfo.org/wf-archive/2002/020314/epf409.htm).

[2] The measures (with sources), "chosen because of the relative quality and objectivity of their data, country coverage, public availability, and correlation with growth and poverty reduction, will be used to assess national performance relative to governing justly, investing in people, and encouraging economic freedom." They are: Governing justly: civil liberties (Freedom House); political rights (Freedom House); voice and accountability (World Bank Institute); government effectiveness (World Bank Institute); rule of law (World Bank Institute); control of corruption (World Bank Institute).

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); three-year budget deficit (IMF/national sources); trade policy (Heritage Foundation); regulatory quality (World Bank Institute); days to start a business (World Bank) (available at http://www.whitehouse.gov/infocus/developingnations/millennium. html).

[3] Exceptions will be allowed by recommendation of the MCA board of directors to the President. Once chosen, recipient countries will sign three-year contracts with the United States, and the effectiveness of their efforts will be judged by the results.

In both scale and design, the MCA has been called the first major foreign aid initiative in more than 40 years. Its underlying logic is that aid can help countries with good governance but will make little difference in countries with bad governance.[4] This is a view expressed in developing countries themselves. For example, the New Economic Partnership for African Development (NEPAD), originated by four African presidents, defines improvements in governance as essential for economic development (United Nations Economic Commission for Africa, 2002; Zirimwabagabo, 2002). For a Latin American example, consider the remarks of Jorge Castañeda, the former Foreign Minister of Mexico:

---

[4] See, for example, Dollar and Pritchett, 1998; Dollar and Kraay, 2000; and Easterly, Levine, and Roodman, 2004. The U.S. Agency for International Development (USAID) has translated these insights into policy pronouncements:

> When development and governance fail in a country, the consequences engulf entire regions and leap across the world. Terrorism, political violence, civil wars, organized crime, drug trafficking, infectious diseases, environmental crises, refugee flows, and mass migration cascade across the borders of weak states more destructively than ever before. They endanger the security and well-being of all Americans. . . . Indeed, these unconventional threats may pose the greatest challenge to the national interest in the coming decades (USAID, 2002, p. 1).
>
> For the past several decades the conventional and, until recently, the predominant perspective on development in the international donor community has been that countries are poor because they lack resources, infrastructure, education, and opportunity. By this logic, if rich countries and international institutions could only transfer enough resources and technology, improve human capacity enough, and support health and education enough, development would occur. To be sure, greater public resources, better physical infrastructure, and stronger public health and education are essential for development. But they are not enough, and they are not the most crucial factor.
>
> No amount of resources transferred or infrastructure built can compensate for or survive bad governance. Predatory, corrupt, wasteful, abusive, tyrannical, incompetent governance is the bane of development. Where governance is endemically bad, rulers do not use public resources effectively to generate public goods and thus improve the productivity and well-being of their society. Instead, they appropriate these goods for themselves, their families, their parties, and their cronies. Unless we improve governance, we cannot foster development (USAID, 2002, p. 33).
>
> Only if governance becomes more democratic and accountable will development occur in the poorly performing countries. And only with a comprehensive, consistent "tough love" from the international community is political will for governance reform likely to emerge and be sustained (USAID, 2002, p. 51).

> For a long period, authoritarian regimes were disguised as presidential ones, states of order were disguised in states of rights, imposing one group's will onto another was disguised under consensus, perpetuating oligarchies were disguised in regimes of altering and the semi-colonial foreign presence and penetration disguised in legal defense of sovereignty (Castañeda, 2003).

Better governance is the key to the solution of the economic problems, adds Castañeda: "This for one simple reason: the only way to pursue structural reforms—if this is the goal to achieve—or to impose a human face to neoliberalism—if this is what is wanted—or to build an alternative to the Washington Consensus—if this is what one wishes—is through institutions which are both democratic and functional, something which Latin America, with rare exceptions, has never benefited from and that is urgent to build" (Castañeda, 2003).

What will the effects be if we choose one or another set of performance criteria for selecting the countries that will receive additional aid?

Abstracting the Problem

This sort of question is not confined to foreign aid. When federal or state governments stress accountability in schools, they hope that rewarding some schools for good performance will create better incentives for districts, schools, teachers, and students. In Pennsylvania, for example,

> The Performance-Incentive Grant Program was created in 1997 to reward individual schools that improve on their own past performance in two primary categories: achievement and effort. Improvement in student achievement is determined from the Pennsylvania System of School Assessment (PSSA) reading and mathematics scores and job-related placements (for Area Vocational Technical Schools), while improvement in effort is measured by increases in student attendance rates. Consistently high-performing schools also are eligible for awards. . . . Recognized schools may use the monies for a wide variety of purposes, selecting those best suited to meet their particular school's educational needs (Pennsylvania Department of Education, 2002).

Allocation by performance measures may mean giving funds to some schools above a threshold of performance but not to those below, as was done, for example, in Florida in the 1980s (Darling-Hammond and Berry, 1988, pp. 51–68).

An abstract version of the problem posits three stages: a legislature (funder) provides money to an executive, who then dispenses the money across recipients (activities, agents). The legislature decides the budget, and the executive decides on the allocation criteria according to measures of performance among recipients.

The legislature maximizes a utility function that is a function of

1. Some results among the agents (Y) and some other dimension of agents' behavior or results (g).
2. The allocation formula eventually chosen by the executive (which may matter apart from 1, say, as a signal of good management) (k = k(g)).
3. The size of the budget (k).

See Table 14.1 for examples.

For simplicity, consider g to be a performance measure that is an imperfect predictor of the value of k in terms of later Y: dY/dk is a positive function of g, with an error term.[5] We might say that Y is produced through an interaction of g and k, along with other factors and with error. We expect that there will be diminishing returns to k to a given agent—otherwise, the executive would give all the aid to the agent with the highest g.

Let us assume the position of the executive. Our task is to choose a set of recipients based on a performance measure g such as to maximize U(Y, g, k), given how agents and our legislature (funder)

---

[5] The legislature, the executive, and the recipient may also value g for its own sake, in addition to g's (imperfect) effect on (later) Y. In the governance example, g might be a measure of democracy, which might be valued as an objective in its own right, as well as an imperfect long-run facilitator of economic development. Similarly, in an educational example, the legislature, the executive, and the schools may all value academic learning for its own sake, as well as for its role as an imperfect contributor to longer-term economic advance, equity, and political development.

Table 14.1
Examples of Choosing and Using Performance Measures

Funding Area	Recipients or Agents	Goal (Y in the future)	Performance Measure or Proxy (g right now)	Investment Based on g (k)
Foreign aid	Countries that receive aid	Economic development	Governance	Foreign aid
Education	Districts or schools	Well-educated citizens	Test scores (or gains in them)	State or federal funding
Health care	Health care providers	Healthy citizens	Proxies for quality of health care, or short-run health indicators	State or federal funding or percentage of reimbursement
A federal agency	Employees	Better public service by employees	Proxies and short-term indicators of employee output	Bonuses, "gain-sharing," and other incentive schemes

will react to our choice and use of g. Our choice and use of performance measures will have several effects:

1. By allocating investments to the agents with higher g, the productivity of k increases in period one. So, compared with equal allocation across agents, there is more Y. The allocation of k that maximizes Y defines efficient static allocation.
2. The selected group may contain a disproportionate number of agents from a particular population of interest. This may lead to accusations of unfairness or bias.
3. Dynamically, agents have an incentive to increase g in order to capture more k in the next round of investment. This has two positive implications: Y will grow more in the future as a function of that higher g, and the marginal impact of each dollar of investment dY/dk in the next round will likely be greater than it was in the first round. But problems may also ensue if agents shift their performance away from productive but unmeasured activities toward measured ones or if they try to "fiddle" the measures of g we use.

4. Our funder may increase (or decrease) our investment budget k in the future because we condition investments on g. The funder may value g for its own sake. The funder may (also) consider that our using the g is evidence that the investments made in the agents will not be wasted. And returning to point 2 above, the funder may react to "inequities" across those selected and those not selected to receive aid.

Analysis for the Case of Foreign Aid

We present our analysis of the use of performance criteria for foreign aid in five parts. First, what measures of governance performance exist, and what are their statistical properties?

Second, how can we analyze efficient static allocation of aid, depending on the way performance measures are used to select a few among many countries for the extra help? If we use g to choose a subset of countries, how much gain in Y can we expect compared with, say, randomly choosing the subset of countries?

Third, how can we examine the representation of groups (of countries, in this case)? How can we portray the tradeoffs of including more members of underrepresented groups?

Fourth, how can we analyze the incentive effects on countries of allocating aid according to governance?

Finally, how can we analyze fundraising effects? Does evidence suggest that aid donors will give more when countries improve their governance?

Measuring Governance

Governance is a popular term, yet defining it is not easy.[6] The term is applied to corporations, universities, and civic associations; in this

---

[6] The Oxford English Dictionary (2d ed., 1989) defines *governance* as 1. The action or manner of governing. b. Controlling, directing, or regulating influence; control, sway, mastery. To govern: 1. *trans.* To rule with authority, esp. with the authority of a sovereign; to direct and control the actions and affairs of (a people, a state or its members), whether

chapter, we concentrate on governments. Most usages include such aspects as popular sovereignty, the size of government, the efficiency of government, the inclusiveness of political and administrative processes, and sometimes political stability. "Good governance" is democratic, limited, efficient, little affected by corruption, open to all members of the population, and stable. "Bad governance" is dictatorial and arbitrary, sweeping in its powers, inefficient, highly corrupt, closed to all but a privileged few, and unstable.

We have collected what we believe are all the publicly available measures of governance, some 40 in all.[7] They differ in coverage, concept, source, and clarity. Little information is available about their reliability or validity, as some scholars have lamented for years (for example, Bollen, 1991, and Inkeles, 1991). We also find little concern with how these different kinds of information might be used together.[8] As a result, we see a phenomenon experienced in many "new areas" of the social sciences: an explosion of measures, with little progress toward theoretical clarity or practical utility.[9]

Confronted with a multitude of possible performance measures, how should we proceed? One question is, How closely related are the various measures? If we have one of them, do we in effect have all of them? Or are they measuring quite different things? After many statistical explorations, including considerable attention to long-tailed distributions, outliers, and bimodality, we find that the many measures of governance are correlated across all the countries in the world in the 0.6-to-0.9 range, with the exception of several of the newer, so-called second-generation governance measures. We also examined the

---

despotically or constitutionally; to rule or regulate the affairs of (a body of men, corporation).

[7] A full description is available from the senior author of this chapter, Robert_ Klitgaard (gaard@prgs.edu).

[8] An exception is the work of Daniel Kaufmann and his colleagues (1999a,b, 2002).

[9] This phenomenon finds a parallel in the early stages of work on measures of personality. Large numbers of psychologists developed their own, relatively untheorized measures of this-or-that angle of personality, tried the measures out on batches of their students, and published the results and the instrument. Only later did other scholars examine the validity, reliability, and interrelationships among the various measures.

correlations among the six composite variables[10] derived in the best data reduction exercise to date (Kaufmann et al., 1999a,b). Using data from 2001–2002, we found that the bivariate correlations among the six composite variables ranged from 0.73 to 0.92. For example, "government effectiveness" and "control of corruption" have a correlation of 0.89. As another example, the correlation between two rival indices of country competitiveness in the World Economic Forum's annual Global Competitiveness Survey—the Growth Competitiveness Index originally developed by Jeffrey Sachs and John MacArthur and the Business Competitiveness Index pioneered by Michael Porter—is above 0.9.[11] Using data through the early 1990s, Klitgaard and Fedderke (1995) found correlations exceeding 0.8 between measures of democracy and measures of corruption.

How might we interpret these correlations? These are imprecise measures. Each suffers from (unknown) measurement error. For normally distributed data, the observed correlation between two variables is equal to the "true" correlation between such variables if perfectly measured times the square root of the product of the reliability coefficients for each variable. Suppose two variables are each measured with a reliability of 0.8, and we observe a correlation of 0.6 between them. Our best guess of the "true" correlation is the observed correlation divided by the square root of the product of the reliability coefficients, or 0.6/0.8 = 0.75. For many social data, reliability is not above 0.8 to 0.9. Thus observed correlation coefficients of 0.6 to 0.8 are high, given the unreliability of measurement. Putting it another way, we would be hard-pressed to say that these highly correlated variables are measuring very different things.

Two of these variables are available for many countries over a long time period (1972 to today): the measures of political rights and of civil liberties developed by Raymond Gastil and now continued by

---

[10] The six composite variables are voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption.

[11] World Economic Forum, 2004. The correlation is from the 2001–2002 report.

Freedom House. These measures correlate between 0.55 and 0.92 with the six composite measures of Kaufmann et al. The canonical correlation between the six Kaufmann measures and the two Gastil measures is 0.95. In our analysis of the broader set of 40 governance measures, we transformed many of the variables to prepare them for factor analysis. In these analyses, a single factor consistently explained "most" of the variance, and the two Gastil measures consistently "correlated highly" with this factor.

However, among developing countries only—a narrower sample—the two Gastil measures do not correlate as highly with important governance variables such as corruption. Among developing countries, the two governance variables based on data from a long period of time—political rights and civil liberties—are correlated with but do not fully capture variables related to the rule of law or the prevalence of corruption. For example, consider the 16 countries selected in May 2004 in the first round of the MCA. Recall that this selection was made on the basis of 16 different variables, including political rights and civil liberties (see footnote 2). If we rank eligible developing countries on the basis of only these last two variables, the top 22 countries—i.e., those with scores of 6 or less on political rights plus civil liberties—include 13 of the 16 countries chosen under the MCA. But they also include nine countries not chosen by the MCA.[12] The agreement is not perfect. So, to check our results, we report below the results of additional statistical analyses that include a larger set of governance variables but over a shorter time period (necessarily so, because of data limitations).

---

[12] The 16 countries selected in May 2004 under the MCA are Benin, Cape Verde, Ghana, Lesotho, Madagascar, Mali, Mozambique, and Senegal in Africa; Mongolia, Sri Lanka, and Vanuatu in Asia and the Pacific; Bolivia, Honduras, and Nicaragua in Latin America; and Armenia and Georgia in Eastern Europe and Central Asia. If we use only the sum of the two variables (political rights plus civil liberties) among countries eligible for the MCA, the 21 best-governed countries (with combined scores of 6 or under) would be found to be these 16 minus Mozambique, Armenia, and Georgia. The best-governed 21 would also include Kenya, India, Kiribati, Papua New Guinea, São Tomé e Principe, Solomon Islands, Guyana, and Albania.

Rewording these results for other performance indicators: First, we have been considering a case where we do not have a strong theory on which to define *performance*, so we have to proceed empirically and examine carefully many possible measures. In the process, we take account of outliers, long-tailed distributions, bimodality, measurement error, and other troubling features of the data.

Second, factor analysis and other multivariate techniques can be useful for determining which measures agree how well, and for exploring whether "performance" appears to be multidimensional.

Third, when one factor captures most of the variance, we may wish to select a few measures that correlate highly with that factor and are widely available.

Finally, if we do choose just a few measures, we sacrifice information. We should examine how some performers deviate from the rest of the population along certain dimensions. And we should compare our results with those obtained using a broader set of measures.

Allocative Efficiency

Once we have tentatively chosen measures of performance based on their theoretical and statistical properties, we turn to the question of their use. Our analysis has four parts: allocative efficiency, representation of groups, incentive effects, and fundraising effects. In this section, we consider the first part. How might we analyze the static efficiency of choosing some countries and not others to receive additional aid?

Fedderke and Klitgaard (1998) showed that various development outcomes and various governance measures go together—although in light of undertheorized models and scant data, it was impossible to establish causality. Barro and Sala-i-Martin (2003, Chap. 12) found that across all countries (not just poor ones), Gastil's two measures have a weak, perhaps curvilinear relationship growth when many other economically relevant variables are taken into account (middling democracies have slightly higher growth than very strong or very weak democracies). They found that a measure of rule of law is positively associated with growth, other things equal (their rule-of-law measure is not publicly available).

In background work for this chapter, we reexamined the relationship between governance and growth (this work will be fully reported in future publications) and found that countries differ. Cointegration analysis of time series for growth and Gastil's governance measures revealed quite different patterns of relationships across developing countries. Thus we cannot readily assume that the relationship between governance and growth is the same in all developing countries.

Using new techniques of panel data analysis across countries, we discovered a useful stratification of the data: For developing countries with ratings higher than 11 on a sum of political rights and civil liberties—in other words, countries with poor governance—we found that investment was lower and the marginal product of each dollar invested was also lower, compared with those of countries with ratings below 10. Our findings supplement the literature. The so-called growth competitiveness index developed by Sachs for the World Economic Forum selects the governance indicators with the highest correlations with growth, holding constant a few other variables (World Economic Forum, 2004). In contrast, we address a different question. In addition to allowing a direct impact of governance on output, we also allow for the possibility of an impact of governance on the level of investment as well as the marginal product of investment.

We examined a population of 66 developing countries from 1972 to 2000. (A number of countries from our earlier analysis of governance measures had to be excluded for lack of data about investment.) The pooled mean group (PMG) estimator we employed exploits the improved power characteristics of a panel by imposing a homogeneous long-run equilibrium relationship across all countries constituting the panel, while allowing for heterogeneity in the dynamics of the specification, as well as fixed effects.[13] Accordingly, we

[13] Note that the solution to the implied difference equation for each country can imply a quite distinct steady state. The advantage of the PMG estimator is that it has greater efficiency than estimators that allow for greater heterogeneity in the panel (e.g., the mean group estimator). Estimators that impose excessive homogeneity on the panel (say, by imposing

test for the presence of long-run homogeneity by means of a Hausman test.[14]

In estimation, we are explicit in recognizing the possible existence of nonlinearities in the association between governance and output, through the possibility of an impact of governance on both the level of investment and the marginal product of capital.[15] Col-

---

homogeneity in both the long run and the dynamics and allowing for only fixed effects, as does the dynamic fixed effects estimator) risk introducing bias and inconsistency in estimation. See the discussion in Pesaran, Shin, and Smith, 1999.

[14] In estimation, we impose a maximum lag length of 3 and choose the lag length for each individual country in the panel by means of an information criterion.

[15] Suppose that

$$Y = f(K,g), \tag{1}$$

such that the level of output depends on a (vector of) standard factors of production such as capital. It also depends on the level of governance. Suppose further that technology has the standard feature that $Y_K>0$, $Y_{KK}<0$, $Y_g>0$, $Y_{gg}<0$. It follows that

$$dY=Y_K dK + Y_g dg, \tag{2}$$

such that output growth depends on governance—or improvements in governance. Estimation of equation (2) subject to an error term may be subject to at least two potential complications. First, accumulation of capital may itself depend on governance:

$$K = K(g), \tag{3}$$

For analytical clarity, assume $K_g>0$, $K_{gg}<0$, such that

$$dY = Y_K dK + (Y_g + Y_K K_g)dg. \tag{4}$$

The impact of any change in governance on output will be both direct ($Y_g dg$) and indirect by altering the *level* of investment ($Y_K K_g dg$). Given the assumptions of $Y_{KK}<0$, $K_{gg}<0$, the impact of changes in governance will be nonlinear in both the level of governance and the level of capital intensity of production. Specifically, at high levels of governance and at high levels of capital accumulation, improvements in governance will have less impact on output than they will where governance or capital stock are low. Second, suppose the marginal product of capital is contingent on the quality of governance. A unit of capital under good governance may contribute more to output than would one under bad governance. Thus we have $Y_K(g)$, and suppose that $Y_{Kg}>0$, $Y_{Kgg}<0$. Then

$$dY = (Y_{Kg} dg + Y_K)dK + Y_g dg. \tag{5}$$

Again the impact of any change in governance will be both direct and indirect—direct through $Y_g dg$, indirect by changing the *impact* any investment has on output, via ($Y_{Kg} dg + Y_K$)dK. Nonlinearity again follows, in this instance across levels of governance. The impact of investment in physical capital on output rises with the level of governance, though at a declining rate. We address these issues through two alternative estimation strategies. Testing for the impact of governance on the marginal product of capital follows immediately by estimating the interaction effect implied by equation (5). Yet this does not serve to identify the

umn (1) of Table 14.2 reports the results from a panel of 66 developing countries.

Columns (2), (3), (6), and (7) report the results for countries with fair or better governance, defined as having scores less than 8 (or less than 10 in columns (3) and (7)) on "governance" (here, a sum of civil liberties and political rights). Columns (4) and (8) give the results for countries with relatively bad governance: Their average scores were above 11. Comparing these columns yields two important findings relevant to the impact of aid.

First, the *impact* on growth of a dollar of investment is higher in countries with good governance than it is in countries with bad governance.[16] Second, results are consistent with a positive association

---

nonlinearity that equation (4) implies. Hence we also proceed by estimating both equations (4) and (5) in a stratified sample of countries: for low, mid-level, and high governance levels. Where governance affects the *level* of investment, we should see statistically significant changes in the coefficient on changing governance. Where governance affects the *impact* of investment, we should see statistically significant changes in the coefficient on investment.

We used data on gross investment in constant 1995 U.S. dollar terms. Strictly, we would like a measure of the change in the capital-labor ratio, given the use of per capita output as the Y variable. However,

$$k = K/L$$

$$dk = (1/L)dK - (K/L^2)dL$$

$$(1/L)dK = dk + (K/L^2)dL.$$

Since $(K/L^2) \rightarrow 0$, it follows that $(1/L)dK \rightarrow dk$; and we are therefore able to estimate dk from the gross investment data modified by population size.

[16] Note that the coefficient on investment in columns (2) through (4) captures the combined effect of the marginal product of investment, as well as the impact of changes in governance on the marginal product of capital (see equation (5) of footnote 15). By contrast, in columns (5) through (8), the coefficient on investment should isolate the marginal product of capital across the groups of countries, while the explicitly included interaction term now identifies the impact of changes in governance on the marginal product of capital. Comparing columns (2) through (4) in the developing countries with fair or better governance ($g<10$ and $g<8$), the impact of investment on growth is statistically significant and two to three times larger than it is in countries with $g>10$. Columns (5) through (8) report findings with an interaction term between governance and investment. With better governance, the marginal product of capital increases, with the impact of investment under sound governance being roughly twice that which holds under poor governance. The interaction term shows that improvements in rights increase the marginal product of investment. What is more, the strongest impact obtains among countries with the worst governance ($g>11$), which have a coefficient roughly ten times as large as that for countries with better governance ($g<8$)—contrast the coefficients for X3 in columns (6) through (8). The efficiency of investment

between better governance and increases in the *level* of investment.[17] Good governance thus appears to bring a double benefit in the form of higher levels as well as higher productivity of investment.

To check these results, we carried out another analysis using a wider range of right-hand-side variables over a shorter time period (because of data limitations over time). When we stratified the countries by measures of governance, once again the impact of investment on changes in per capita income is higher in countries with good governance (better political rights and civil liberties) than in those with bad governance. For countries with good governance, the coefficient on investment is between 0.80 and 0.98, depending on model specification and included covariates, while for countries with poor governance, it is between 0.29 and 0.46. (These results will be reported fully in a future publication.)

These results support the underlying idea of the MCA. In terms of the productivity of additional investment (such as aid), countries with poor governance do seem different from countries with good governance. If one wishes to select some among many developing countries for additional aid and one has the goal of allocating the aid to produce the most growth, a solution is to omit countries with poor governance.

How much additional growth would be obtained by using one or another performance criterion to select the subset of recipient countries? We have been considering here allocative efficiency, without yet taking into account incentive effects. In this vein, one could

improves with governance, with the strongest increase obtaining for moving out of the worst-possible-governance category.

[17] Note that where governance impacts the level of investment, the coefficient of governance captures both its direct marginal impact on output and the indirect impact via changes in the level of physical capital stock (see equation (4) of footnote 15). Given our finding that the marginal product of capital increases with improvement in governance, and presuming standard concavity of output in governance, the expectation is of a decline in the absolute magnitude of the coefficient on governance, if the level of investment rises with improvement in governance, though under strong concavity assumptions. The evidence of both columns (2) through (4) and columns (6) through (8) can be shown to be consistent with this prior.

Table 14.2
Investment, Governance, and Growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**
	Full Sample	**Governance <8**	**Governance <10**	**Governance >11**	**Full Sample**	**Governance <8**	**Governance <10**	**Governance >11**
Info Crit:	AIC(3)	AIC(3)	AIC(3)	AIC(3)	ARDL(3,3,3,3)	ARDL(3,2,0,1)	ARDL(3,2,0,1)	ARDL(3,1,1,3)
Y:	Growth	Growth	Growth	Growth	Growth	Growth	Growth	Growth
X1: Investment	0.005*	0.006*	0.005*	0.002	0.007*	0.007*	0.005*	0.003
	(0.001)†	(0.002)†	(0.002)†	(0.005)†	(0.002)†	(0.002)†	(0.002)†	(0.005)†
X2: dGovernance	0.000	-0.001	-0.001*	-0.010*	0.000	-0.003*	-0.003*	-0.075*
	(0.001)	(0.001)	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)	(0.023)
X3: X1*X2					-0.002*	-0.002*	-0.001*	-0.019*
					(0.001)	(0.001)	(0.001)	(0.007)
N	66	29	43	21	64	29	43	21
δ	-1.03*	-0.97*	-1.05*	-0.82*	-0.85*	-0.81*	-0.93*	-0.77*
	(0.04)	(0.06)	(0.06)	(0.08)	(0.05)	(0.07)	(0.06)	(0.11)
h-test	3.88	1.03	1.10	3.96	2.79	3.85	1.38	X1:1.39[0.24]
	[0.14]	[0.60]	[0.58]	[0.14]	[0.43]	[0.28]	[0.71]	X2:1.51[0.22]
								X3:2.05[0.15]
Constant	0.02*	0.02*	0.02*	0.007*	0.02*	0.02*	0.02*	0.01*
	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
dY(-1)	0.04	0.03	0.06	-0.09	-0.03	-0.07	-0.02	-0.06
	(0.03)	(0.04)	(0.04)	(0.07)	(0.04)	(0.05)	(0.04)	(0.06)
dY(-2)	-0.01	-0.02	-0.02	-0.054	-0.07*	-0.06	-0.04	-0.05
	(0.02)	(0.02)	(0.02)	0.044	(0.03)	(0.04)	(0.03)	(0.06)
dX1	0.14*	0.15*	0.14*	0.12*	0.13*	0.15*	0.15*	0.11*
	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.03)

Table 14.2 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimator	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**	**PMGE**
	Full Sample	**Governance <8**	**Governance <10**	**Governance >11**	**Full Sample**	**Governance <8**	**Governance <10**	**Governance >11**
Info Crit:	AIC(3)	AIC(3)	AIC(3)	AIC(3)	ARDL(3,3,3,3)	ARDL(3,2,0,1)	ARDL(3,2,0,1)	ARDL(3,1,1,3)
Y:	Growth	Growth	Growth	Growth	Growth	Growth	Growth	Growth
dX1(-1)	0.02*	0.01	0.02*	0.03	0.02*	0.01	0.02*	
	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	
dX1(-2)	0.01*	0.01*	0.012*		0.01			
	(0.004)	(0.01)	(0.005)		(0.01)			
dX2	0.001	0.000	0.000	-0.01	0.01			0.07
	(0.001)	(0.002)	(0.001)	(0.01)	(0.02)			(0.04)
dX2(-1)	0.000	0.001	0.000	-0.01	-0.01			
	(0.001)	(0.001)	(0.001)	(0.01)	(0.01)			
dX2(-2)	0.001	0.000	0.000	0.00	0.003			
	(0.001)	(0.001)	(0.001)	(0.01)	(0.01)			
dX3					-0.01	-0.01	-0.005	0.03
					(0.02)	(0.01)	(0.004)	(0.02)
dX3(-1)					-0.01			0.001
					(0.01)			(0.002)
dX3(-2)					0.01			0.000
					(0.01)			(0.001)

NOTE: Y = growth in real per capita GDP. Governance = the sum of the two Gastil measures (political rights + civil liberties, where 2 is best and 14 is worst). Investment = the rate in real physical capital stock.
d = the difference operator. * = significance at the 5% level. ** = significance at the 10% level. † = the variable concerned was under natural logarithmic transform. N = number of countries in group. δ = speed of adjustment to long-run equilibrium. h-test = the Hausman test for long-run homogeneity. Round parentheses denote standard errors; square brackets denote probability levels

carry out simulations of the growth that would follow from using various performance measures to allocate the aid. These simulations would be based on econometric estimates such as those we have been considering: If we use these performance criteria to select the k countries among N possible recipients, the result would be a Y percent increase in growth.[18] Psychometrics provides another method for assessing the efficiency of a selection. For simplicity, suppose we are to select a proportion k/N or π of N countries (and the countries are of equal size). It is proposed that we use g, an indicator of each country's (governance) performance now, which we value solely as a predictor of a valued objective in the future (Y). How much of an increase in Y will we get by selecting π using g?

Applying selection theory under normality, the gain per country turns out to be

$\Delta E(Y)$ per country = $r\sigma_Y\varphi/\pi$.[19]

---

[18] Simulations based on our estimations, accounting for possible nonlinearities between governance and output growth, suggest that other things being equal, growth increases from 0.5 percent per annum to 3 percent per annum as countries move from the worst level of governance (>11) to midrange governance (>7, <10) and that growth then settles down to roughly 1.5 percent for good governance (<7).

[19] For simplicity, assume the data are well-behaved; that g is normalized so it has a mean of 0 and a standard deviation of 1; and that we end up with a partial correlation r of Y and g given k. Suppose we define $\mu(Y)$ as the mean of Y among all developing countries. Then the regression of Y on g (after adjusting for other variables) is

$$Y = \mu Y + \beta g + e, \tag{1}$$

where e is a random error term.

If we select recipient countries *on the basis of g* (not randomly), what is the average Y of the selected group of recipients?

$$E(Y_s) = E(\mu Y) + E(\beta g_s) + E(e), \tag{2}$$

where the subscript s means *in the selected group*. Since $E(e) = 0$ and μY and β are constants, this becomes

$$E(Y_s) = \mu Y + \beta g_s\, E(g_s). \tag{3}$$

Since $\beta = r(\sigma Y/\sigma_g)$, where σY is the standard deviation of Y of all recipients and $\sigma_g = 1$, $\beta = r\sigma Y$. Thus

$$E(Y_s) = \mu Y + r\sigma Y E(g_s). \tag{4}$$

For normally distributed data,

$$E(g_s) = \varphi/\pi, \tag{5}$$

Here $\Delta E(Y)$ is the change in expected Y, r is the correlation we compute between g and Y in the entire sample (not just the π selected), σ_Y is the standard deviation of Y in the population of countries, π is the proportion of countries we wish to select, and φ is the ordinate of the standard normal distribution corresponding to that.[20]

Statistical analyses of this genre could help us estimate the allocative efficiency of different ways to allocate foreign aid. We could examine what might happen to total GDP across all aid recipients if we allocated more aid to countries with good governance and less to countries with bad governance. In addition to the "best guess" about these effects of allocative efficiency, we would report the uncertainties surrounding the predictions.

Representation and "Fairness"

There is a second point in the use of performance measures: group representation. If certain groups differ in their scores on a performance measure, then using that measure to select will lead to an underrepresentation of lower-scoring groups. For example, the MCA will exclude a disproportionate number of African countries (Brainard and Driscoll, 2003). Predictably, this will lead to accusations of "unfairness" to Africa.

The MCA seems to anticipate underrepresentation by degree of poverty. It segments the poor countries into two groups, poor and very poor. Otherwise, on the basis of governance measures, "too few" very poor countries might be selected. Too few in what sense? Per-

---

where φ is the ordinate of the standard normal distribution corresponding to a π probability of being selected.

To compare the expected increase in Y from using g to select π, we would take equation (4) and subtract the expected Y if selection were random, which is μY. Thus, the gain per country is

$$\Delta E(Y) \text{ per country} = r\sigma_Y\varphi/\pi. \quad (6)$$

[20] As an example, suppose we find r = 0.4. Suppose we select one in five countries to receive aid, so $\pi = 0.2$. The normal tables tell us that for $\pi = 0.2$, φ is 0.28. Thus $\varphi/\pi = 1.4$. What about σ_Y? Recall that this is the standard deviation of Y in all countries. If Y is GDP growth and $\sigma_Y = 3$ percent, then the gain per country selected in expected Y is 1.68 percent. Just by selecting countries with better governance, we will end up with countries with higher GDP growth.

haps not in terms of allocative efficiency or incentive effects on the countries, but too few in some dimension of representation or fairness.

The phenomenon of underrepresentation is quite general in selection models and allocation models. Around the world, we are familiar with this problem with regard to personnel selection, merit pay, and university admissions, where the use of merit ratings leads to the underrepresentation of certain disadvantaged groups (Klitgaard, 1986; Klitgaard, 1990, Chaps. 10–12; Sowell, 2004).

Policymakers often face a tradeoff between efficient selection and underrepresentation. The tradeoff depends on value judgments—how much do you value a such-and-such percentage increase of members of group A among those selected? But it also depends on factual matters. How much do you give up in performance to get more members of group X among those selected? Answers can be provided in terms of g and in terms of forgone Y.

The Appendix provides a tool to help decisionmakers understand possible tradeoffs between efficient allocation and group representation. The tradeoffs depend on specific features of the particular selection problem, such as the strength of the predictive relationship r, the proportion π of agents chosen, the differences among groups in the performance measures g, the shares of the various groups among the agents, and the value we give to later outcomes Y.

Incentive Effects

The third dimension of using performance measures concerns the incentives created for recipients (agents, programs). What might recipient countries do if we choose to allocate new aid to countries with good governance?

Milgrom and Roberts studied a general version of this problem. They found that "the strength of incentives should be an increasing function of the marginal returns to the task, the accuracy with which performance is measured, the responsiveness of the agent's efforts to incentives, and the agent's risk tolerance" (Milgrom and Roberts, 1992, p. 240). Transferring this to our problem, how strongly we

should condition aid on performance (governance) is an increasing function of the marginal returns to Y of the recipient's "effort" (for which g is a performance measure), the accuracy with which g is measured, the responsiveness of the countries to the incentives, and countries' attitudes toward risk.

These conditions will vary across countries. One might speculate that the most recalcitrant countries will be those where

- Geography, poor human and physical capital, and instability mean that better governance will have a small payoff in terms of growth.
- Leaders and citizens are so poor that they will resist entering any new aid scheme in which they might lose resources.
- Leaders and perhaps citizens deny the validity of Western concepts and measures of "good governance."
- It is easy to dissimulate good governance or to manipulate the performance measures used.
- Aid is a small part of the recipient's total budget.
- Leaders benefit personally from bad governance.

Thus, in terms of the incentives created, under some conditions a donor should give great weight to governance performance criteria in allocating aid, but under other conditions the best choice is an amount of aid that does not vary with performance. Table 14.3 illustrates some extreme cases.

Note that the incentive effects depend on the particular governance measures we choose. Suppose we have several measures with more-or-less equivalent predictive power. If we choose a measure that is beyond a country's control, then of course it will have no incentive effect (except frustration). For example, Acemoglu, Johnson, and Robinson (2001) found that differences across countries in the extent of property-rights enforcement can explain the bulk of the differences in income per capita. But they argue that the underlying cause is different colonization experiences and that these differences led to different institutional developments that still affect economic out-

Table 14.3
When Recipients Will Respond Positively to Allocating Aid by Governance

Recipient Characteristic	Responsive When	Unresponsive When
Marginal benefit of more effort by recipient on future GDP	Better governance leads to rapid economic growth	Because of other constraints in the country, better governance has little effect on economic growth
How accurately governance predicts future GDP growth	Governance can be measured accurately and cheaply (and without controversy); governance is highly correlated with the country's "development effort"; performance measures cannot easily be dissimulated or manipulated	Governance measures are inaccurate, expensive, and controversial; governance is only weakly correlated with a country's "development effort"; performance measures can easily be dissimulated or manipulated
Responsiveness of recipient's effort to governance-based aid incentives	Recipient is responsive to governance-conditioned aid—perhaps because aid is a large part of the recipient's budget, perhaps because improvements in governance are valued by the recipient	Governance-conditioned aid is a small part of the recipient's budget; improvements in governance are not in the interests of the recipient's leaders
Recipient's risk aversion	Recipient countries are almost risk-neutral	Recipient countries are very risk-averse, perhaps because they are poor

comes. Suppose an unwise reader of their conclusions decided to use colonial heritage as a measure of governance. Since a country today has no control over that variable, using this measure would have no incentive effects. In contrast, a country can affect such measures as political rights, civil liberties, and corruption.

Some measures may be more easily manipulated or dissimulated than other measures. The chapters by Asch, Hamilton, and Klerman in this volume describe how performance measures can be gamed or corrupted.

Finally, incentives are particularly powerful right around the "cut point," where a country is selected or not. If a country is far below the cut point, it may feel little incentive to improve, because it

can't conceive of improving enough to be chosen. Studies of affirmative action have noted this theoretical possibility.

When choosing performance measures, we should take into account a variety of incentive effects as well as allocative efficiency. And a final dimension, how our funders will respond, should also be considered.

Fundraising Effects

Actors besides the donor and recipient are often important in choosing and using performance measures. For example, the aid USAID allocates is part of the State Department's budget, submitted by the President and approved by Congress; in some sense, the budget is ultimately affected by voters' preferences. How well USAID spends the money—the impact the aid has, the accountability USAID demonstrates—influences how much money USAID gets in its next budget.

In our illustrative example, the budget we have is a function of the performance criteria we use for allocation. We call this the *fundraising effect*. This effect can emerge for two reasons: First, the administration, Congress, and the people may value good governance for its own sake, as an objective of aid apart from GDP growth. This is probably especially true for measures of political rights and civil liberties. And second, they may believe that the leakage of aid will be lower if we give it only to countries with good governance. This is probably especially true for measures of rule of law, government efficiency, and (low) corruption.

Fundraising effects arise in other examples of allocation according to performance criteria. For example, if an education agency puts a strong emphasis on allocating educational budgets according to performance on standardized tests, one result may be that the legislature and the people decide to spend more money on education.

This point has been recognized by economists and political scientists in the literature on poverty targeting, but to our knowledge, it has not been explicitly included in analyses of performance-driven allocation systems. Nichols and Zeckhauser (1982) pointed out that food stamps may be more efficient for the poor than theoretically optimal lump-sum transfers if those providing the budget for aid to the

poor value food-based aid. Gelbach and Pritchett (1997) created a model in which the policymaker chooses the performance criteria for aid allocation, but the budget for aid is determined through majority voting. Most voters do not like the idea of a program benefiting only the poor, so the majority oppose targeting aid. If we ignore political feasibility and assume that the budget is fixed, we will choose full targeting of transfers. But in response to this choice of "performance"-based allocation, in the Gelbach/Pritchett model, the legislature reduces the budget, and consequently the poor receive little. In contrast, when we recognize budgetary endogeneity, we give aid to everyone, and the aid budget grows. The poor actually do better under this scheme than they do under an allocation formula that seemingly favors them.

How does this analysis apply to foreign aid? Note that it undercuts critiques of tied aid (aid that a country insists its nationals provide, even if nonnationals can provide the good or service more cheaply). Critics point out that untied aid gives a recipient access to lower prices and higher quality through an unrestricted market, and some estimates put the gains at 20 percent of the aid received. But the critics ignore the likelihood that untied aid would win fewer votes in Congress. The aid budget might plummet, leading to fewer goods and services being available to recipients.

For the case of governance and foreign aid, we wish to know how the State Department, the administration, Congress, and the citizens of the United States will respond to the MCA. Will the conditioning of additional aid on governance lead to more support for this aid? Put another way, if the MCA's conditioning on governance were scrapped, would the additional aid be scrapped as well? The answers go beyond the bounds of this chapter. But in the analytical spirit of this discussion, it is useful to consider a related question and to consider what one can infer from historical data. Do donor countries and international financial institutions give more aid to countries with better governance, other things being equal?

We have analyzed bilateral aid flows from 1975 to 1999 from 21 donor countries (including the United States) to 144 recipient countries, using two dependent variables: the chance a country would

receive aid and the amount of aid it would receive. The independent variables included the country's governance[21] and a variety of other factors, including population, GDP p.c., colonial ties to the specific donor, continent, aid from other countries (to measure a "bandwagon effect"), and trade flows to the donor relative to GDP. Akramov used a variety of estimation techniques and specifications.

We found that in most donor countries, the quality of a recipient's governance has *not* been an important driver of foreign aid decisions. In only four of the 21 donor countries (Canada, Denmark, Sweden, and the United States) does the governance variable have a statistically significant positive effect on the probability of giving foreign aid. In only three of the donor countries (Belgium, Germany, and New Zealand) does the governance variable have a statistically significant positive effect on per capita aid flows—and again, the effect is only mildly important. These results suggest that only seven bilateral donors seem to reward good governance one or another way (however, those seven donors provide about 46 percent of total bilateral foreign aid).

The case of the United States is of course of most relevance to this chapter. The analysis suggests that a country moving from the mean level of governance among recipients to one standard deviation above the mean raises the probability that the United States will give that country aid by 71 percent. But once the United States decides to give a country aid, the amount is not a function of the recipient's governance.

What can we conclude from this historical analysis for current U.S. policy regarding the MCA? Especially compared with other donors, the United States has for the past quarter-century already been giving governance considerable implicit weight in selecting the recipients of bilateral aid. The MCA is new in many ways, but the United States has already been selecting recipients on the basis of political rights and civil liberties.

---

[21] Governance was measured using a combination of g1 and g2 (political rights and civil liberties) constructed through a canonical correlation with the six Kaufman et al. composite measures.

What can we infer about the likely behavior of American aid in the future? Arguably, not much, because times have changed, in the United States and elsewhere. To predict the fundraising effects of the MCA's emphasis on good governance, we would like to know how much Congress' willingness to fund a 50 percent increase in aid depends on the use of "tough love." Without thresholds in areas such as corruption and democracy, would Congress be likely to agree to the MCA, or might the agreed-upon budget be much smaller?

By announcing our selection criteria, we send several kinds of signals and create several kinds of incentives. We signal a policy of broader import in our administration ("we allocate by results"). We signal to our international partners (donors, recipients) our values of democracy and good government, which matter, for example, in foreign policy. We support subsets of countries that are undertaking governance reforms, including NEPAD.

How large will these effects be? These results cannot tell us. The point in doing the econometric work is rather a methodological one, perhaps applicable for other contexts. Sometimes we are looking at problems in which times have not changed, and we can examine the apparent preferences of the legislature (the funder) over time. This will give us ideas about whether our use of performance mea-sures might lead to an increase (or a decrease) in funding for the next period. We quickly learn, however, that estimating the funder's decisions is not easy, either theoretically (many factors matter) or empirically (ideally, we would need data from many time periods). But econometric modeling can help as an analytical guide to the questions we should be asking directly of the funder(s).

Implications for Choosing and Using Performance Measures

The Complexities of Aid Policies

As we turn to policy implications, we must emphasize the limited scope of the foregoing discussion of foreign aid. We have been analyzing development assistance, which is conventionally separated

from humanitarian or relief aid, from military assistance, from private philanthropic activity, and from commercial credits. In the real world, the separation is not stark. Military assistance has developmental impacts, not always good ones; so do food aid, disaster relief, and export or investment credits.

Even with development assistance, our country's objectives are numerous and complicated. The legislature and the executive want to increase growth, reduce poverty, enhance human rights and dignity, protect vulnerable groups and cultures, prevent illegal migration, strengthen democracy, reduce global warming, improve international understanding, and create nations capable of resisting terrorism. At last count, the U.S. Foreign Assistance Act of 1961 as amended posits 33 different development goals and 75 priority areas; each USAID project has to say what it will do for the environment, women, children, and so forth. *Development* means many things, not necessarily tightly connected and not necessarily agreed upon among "us" or "them."

For these things, the relevant utility functions are surely nonlinear. We value a $100 increase in average annual income much more between $200 and $300 than between $3,000 and $3,100. The utility functions may also have national and continental subscripts. For example, even if it were possible to reduce more poverty by focusing only on India and Nigeria (say), we might want to make sure that at least some aid is going to every continent.

As the United States allocates foreign assistance, it has always recognized that aid is (or should be) more than simply money. When the United States decides how much to give to whom, it should also ask *what* it should give—for example, technical assistance. What is the United States especially good at providing, compared with the recipient's needs and with other donors' capabilities? And *how* should the aid be provided? The United States may condition aid on actions by the recipient: "You only get the aid if you do this and that." Conditionality may offend the recipient as interference. On the other hand, conditionality is well known in the private sector, where venture capitalists may invest only if certain conditions are met, even a requirement to give the investor a seat on the board of directors.

Sometimes conditionality has been welcomed by the recipient, even if not publicly.[22]

So aid policy is more complex than our simple allocation problem above (see Table 14.4). With aid, we are trying to achieve many objectives, only one of which is growth. We weight growth by a country's level of income and perhaps by other geographical factors. We are giving not just money, but factors of production (technologies, knowledge, skill) that may have a value in the country much greater than our cost or than a certain amount of money (because the country would allocate money in other ways). How we give the aid matters, for example, in the conditions we attach that help or hinder local commitment.

Implications

In this chapter, we have forgone the complications of aid policy to emphasize points sometimes overlooked in discussions of choosing and using performance measures.

Table 14.4
Reality Check

Characteristic of Aid	Simple Allocation Problem	Real-World Problem
Objective	Maximize recipient income	Complicated by multiple developmental objectives and nonlinear utility functions
For whom	All countries (all poor countries)	Country and regional subscripts may matter
What is given	Money	Skill, technology, knowledge, and other things, meaning that the donor's comparative advantage matters
How it is given	A gift (a check)	Conditional assistance, perhaps in the form of a project or a contract

[22] In the 1960s, Peruvian President Fernando Belaúnde Terry wanted land reform but was blocked by the oligarchs in parliament. The Alliance for Progress worked with Belaúnde Terry to create a set of conditions for Peru to receive the alliance's aid. One of the conditions was land reform. In public and in parliament, Belaúnde Terry protested mightily against this condition. But in private he welcomed it, as it enabled him to win a bargaining game with his own parliament.

First, as in our problem, performance is often "undertheorized" in many areas of public policy, meaning that we have no agreed-upon set of measures to use for selecting recipient countries—or in related problems, for selecting schools, health programs, or employees for special funding. Statistical analysis can help clarify the relationships among proposed indicators of performance. We have learned that across all countries, most measures of g are highly correlated. In our main econometric work, we used two among 40 possible governance measures because those two were correlated with the others and because data about the two were available for many years and many countries. Nonetheless, some countries and groups of countries may perform differently on different measures, so using only a few measures is an imperfect convenience.

Second, we have shown how heterogeneity may matter—and how it can be analyzed. Heterogeneity here implies that the relationship between the performance measure now and what we value later—between governance and later GDP p.c.—is not the same across countries. We used cointegration analysis for individual developing countries to show that our governance measures and GDP p.c. do not go together in the same ways across countries. Then, in panel data analysis, we employed a method that allows for heterogeneity and discovered that countries with "worse governance" are different from countries with "better governance." The results enable us to get a better idea of what might happen if aid recipients were selected according to their governance. Along the way, we separated two issues that are often confused: the marginal effect of additional aid on Y given g, which is not the same thing as the marginal effect of an increase in g on Y. For other examples of choosing and using performance measures, these same lessons may be important: pay attention to heterogeneity, use estimation techniques that take it into account, and focus on the right question of allocative efficiency (the productivity of the additional funds in terms of Y, given each recipient's performance measure g).

Third, the choice of performance measures may have effects on the representation of certain groups of recipients among those selected. If we select countries on the basis of their governance, we get a

group in which investment is more productive. However, we also may get too few countries from certain groups of interest, such as the very poorest countries, or from a particular region. We may want to define our selection procedure to give weight to such factors or to stratify the selection. In the case of the MCA, it is likely that African countries will be underrepresented among those that meet the governance criteria for selection. How we might trade off efficiency and representation is the subject of the Appendix.

Fourth, allocative efficiency is not all that matters when we choose and use performance measures. We should also care about the incentives our performance measures create. A simple model here yields interesting qualitative results that seem to have general applicability. The power of the incentives we create for recipients depends on four factors: the marginal returns of the agent's "effort" to future output, the accuracy with which effort is measured by our performance measures, the responsiveness of the agent's effort to incentives, and the agent's risk tolerance. Agents will change their behavior as the result of performance incentives when

- Their efforts to improve performance measures have a significant effect on valued outcomes in the future.
- Agents are not so risk-averse that they will resist entering any new incentive scheme where they might lose resources.
- Given their own objectives, agents accept the validity of the performance measures.
- It is not easy to dissimulate good performance or to manipulate the performance measures used.
- The performance incentive is a significant proportion of an agent's total budget or paycheck.
- Agents do not benefit personally from bad performance (e.g., via corruption).

We can translate these ideas to other domains besides foreign aid, such as schools, health programs, and federal employees.

Fifth, the performance measures we choose have fundraising effects. Our funders may approve (or disapprove) of the measures we

use. We have considered the perhaps surprising theoretical result that targeting aid in certain ways may lead to a reduction in the budget due to voter response; this may leave those whom we target worse off than if we didn't target at all. More to the point of the foreign aid problem, there is historical evidence that the United States has chosen aid recipients on the basis of political rights and civil liberties. One might conjecture that the MCA will enable an increase in aid by extending and formalizing this historical trend. But this is only conjecture; more generally, we emphasize that the choice of performance measures should consider the effects not only on recipients but also on our sources of money—thus the moniker, *fundraising effects*.

How might all these considerations be taken into account? They are still not sufficient for the complications of the foreign aid problem, as we have seen; data are too scarce, theory is too weak, and complications are too many. And yet, even in their relative simplicity, combining them exceeds the powers of statistical analysis and formal modeling.

Despite their limitations, the factors we have presented can still be used to guide discussion among policymakers, legislators, and recipients. The following questions might be asked when choosing and using performance measures:

- What performance measures could we use to select those who receive additional help or incentives? What are the statistical properties of these measures? Do they tend to go together? Do they cluster in certain groups?
- How well do various performance measures predict more ultimate outcomes that we value? How does using one set of measures or another affect the impact of an additional dollar of investment? Do these predictive relationships vary across agents or groups of agents? If so, can we take heterogeneity into account when we assess the value of using one performance scheme or another?
- Will some groups of recipients be underrepresented among those selected? How might we trade off efficiency and representation in the way we use performance measures?

- What incentives are created for recipients by the performance measures we choose and how we use them? How might we enhance the good incentives and dampen the nefarious ones?
- What will be the reactions of donors and recipients to the performance measures we choose and the way we use them?

The models presented here provide ways to get rough estimates of possible results—methods that can also be used to encourage a dialogue among policymakers, recipients, and legislators. We hope this combination of analysis and dialogue will improve the ways we choose and use performance measures, both in foreign aid and more generally.

APPENDIX

A Model for Trading Off Efficiency and Representation in Selection[23]

The indented boldface sentences are *Mathematica* commands. The indented plain-face sentences and the graphics are *Mathematica* output. The particular parameters chosen are for illustrative purposes only. Read in graphics and statistics packages if needed.

```
Needs["Statistics'ContinuousDistributions'"]

Needs["Graphics'ImplicitPlot'"]
```

Define the probability density functions f_A and f_B of the performance measure for groups A and B. A performance measure could be a test score, such as a SAT or GRE score, or a measure of productivity, such as citations per time period.

```
f_A[t_] := PDF[NormalDistribution[μ_A, σ_A], t]

f_B[t_] := PDF[NormalDistribution[μ_B, σ_B], t]
```

Define the cumulative distribution functions F_A and F_B of the performance measure for groups A and B.

```
F_A[t_] := CDF[NormalDistribution[μ_A, σ_A], t]

F_B[t_] := CDF[NormalDistribution[μ_B, σ_B], t]
```

[23] The *Mathematica* code for this appendix was prepared by Prof. Michael Mattock of the Pardee RAND Graduate School.

Define functions g_A and g_B that transform performance measures into outcome measures. For example, a function could transform a SAT score into expected college GPA.

```
gA[x_] := αA + βA x

gB[x_] := αB + βB x
```

Define functions h_A and h_B that are the inverses of g_A and g_B.

```
hA[o_] = x/ . Solve [gA[x] == o, x] [[1]] // Simplify
```

$$\frac{o - \alpha_A}{\beta_A}$$

```
hB[o_] = x/ . Solve [gB[x] == o, x] [[1]] // Simplify
```

$$\frac{o - \alpha_B}{\beta_B}$$

Define a group of example parameters; n_A and n_B are the numbers of applicants in groups A and B, respectively.

```
Parameters =

{ μA → 50, σA → 10, αA → 0, βA → 1/2, μB → 60, σB → 10,
  αB → 0, βB → 3/4, nA → 20, nB → 80 }
```

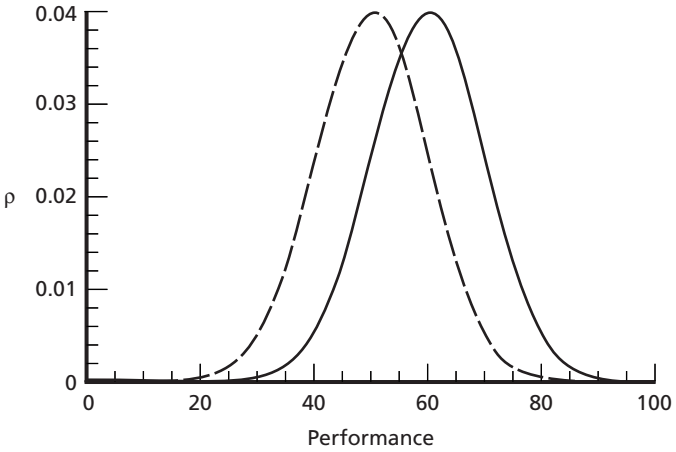
$$\left\{\mu_A \to 50, \sigma_A \to 10, \alpha_A \to 0, \beta_A \to \frac{1}{2}, \mu_B \to 60, \sigma_B \to 10, \alpha_B \to 0, \beta_B \to \frac{3}{4}, n_A \to 20, n_B \to 80\right\}$$

Plot the distributions of the performance measure for the two groups. The dashed line represents group A, and the solid line represents group B.[24]

---

[24] The Mathematica instructions are written to generate output in color, as indicated by RGB in the code. For this volume, we have substituted line styles for color.

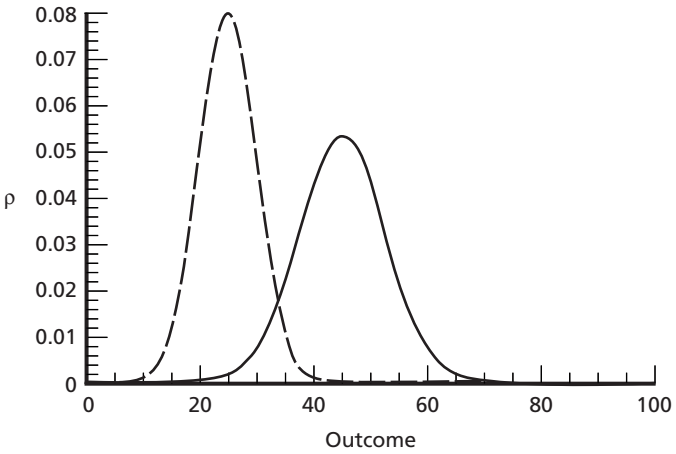
```
Plot[Evaluate [{f_A[t], f_B[t]} / . Parameters],
  {t, 0, 100}, PlotStyle → {{RGBColor [1, 0, 0], Thickness [0.01]},
    {RGBColor [0, 0, 1], Thickness [0.01]}},
      AxesLabel – > {Performance, ρ}]
```



Plot the distributions of the outcome measure for the two groups. The dashed line represents group A, and the solid line represents group B.

$$\text{Plot[Evaluate}\left[\left\{\frac{f_A[h_A[t]]}{\beta_A}, \frac{f_B[h_B[t]]}{\beta_B}\right\} / . \text{ Parameters}\right],$$

```
  {t, 0, 100}, PlotStyle → {{RGBColor [1, 0, 0], Thickness [0.01]},
    {RGBColor [0, 0, 1], Thickness [0.01]}}, AxesLabel – >
      {Outcome, ρ}, PlotRange – > {0 , 0.08}]
```



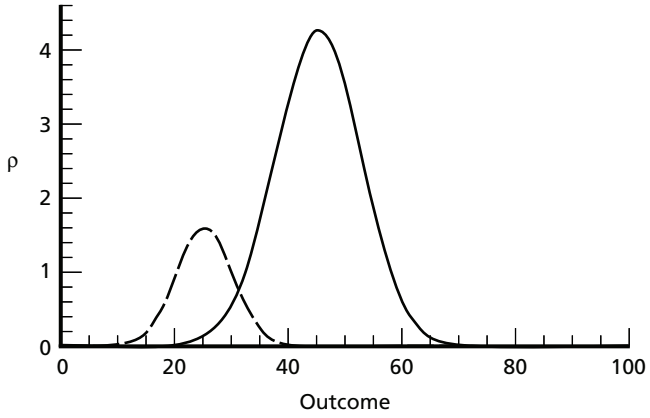
Plot the distributions of the outcomes for the two groups according to their relative sizes. The dashed line represents group A, and the solid line represents group B.

$$\textbf{Plot[Evaluate}\left[\left\{n_A\,\frac{f_A[h_A[t]]}{\beta_A}\,,\,n_B\,\frac{f_B[h_B[t]]}{\beta_B}\right\}\textbf{/ . Parameters}\right],$$

{t, 0, 100}, PlotStyle → {{RGBColor [1, 0, 0],

Thickness [0.01]}, {RGBColor [0, 0, 1], Thickness [0.01]}},

PlotRange – > {0 , 4.5}, AxesLabel – > {Outcome, ρ}]



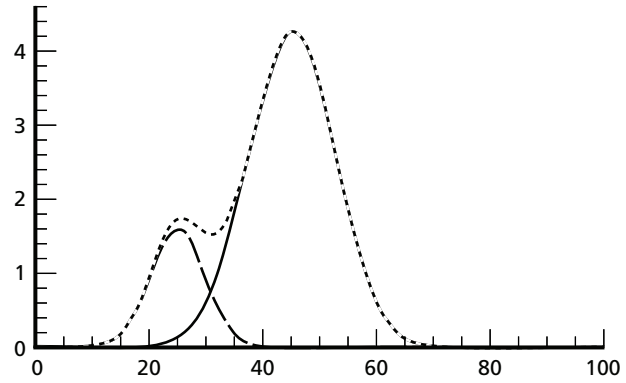
Define f, the combined probability density function over outcomes for the entire population.

$$f[t\_] = \frac{n_A \frac{f_A[h_A[t]]}{\beta_A}}{n_A + n_B} + \frac{n_B \frac{f_B[h_B[t]]}{\beta_B}}{n_A + n_B} \text{ // Simplify}$$

$$\frac{E^{-\frac{(-t+\alpha_B+\beta_B\mu_B)^2}{2\beta_B^2\sigma_B^2}} n_B\beta_A\sigma_A + E^{-\frac{(-t+\alpha_A+\beta_A\mu_A)^2}{2\beta_A^2\sigma_A^2}} n_A\beta_B\sigma_B}{\sqrt{2\pi}(n_A + n_B)\beta_A\beta_B\sigma_A\sigma_B}$$

Plot the distributions according to their relative sizes. The dotted line represents the total population distribution, while the dashed line represents group A, and the solid line represents group B.

```
Plot[Evaluate
  [{n_A f_A[h_A[t]]/β_A, n_B f_B[h_B[t]]/β_B, (n_A + n_B)f[t]} /. Parameters],
  {t, 0, 100}, PlotStyle → {{RGBColor [1, 0, 0],
    Thickness [0.01]}, {RGBColor[0, 0, 1], Thickness[0 .01]},
    {RGBColor [0, 1, 0 ], [Thickness  [0.01 ]}}, PlotRange –> {0, 4.5}]
```



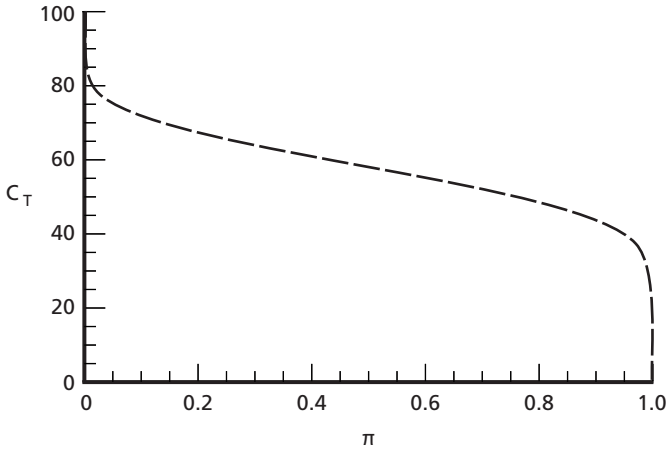
Define F, the combined cumulative distribution function for the performance measure for the entire population.

$$F[b\_] = \frac{n_A F_A[b]}{n_A + n_B} + \frac{n_B F_B[b]}{n_A + n_B} \text{ // Simplify}$$

$$\frac{\left(1 + \text{Erf}\left[\frac{b - \mu_A}{\sqrt{2}\,\sigma_A}\right]\right) n_A + \left(1 + \text{Erf}\left[\frac{b - \mu_B}{\sqrt{2}\,\sigma_B}\right]\right) n_B}{2(n_A + n_B)}$$

Plot the cut score (ignoring representation) C_T versus π, the fraction of the total applicant population to be selected.

```
ParametricPlot[{(1 – F[t] / . Parameters), t}, { t, 0, 100 } ,
  AxesLabel – > {π, C_T}, AspectRatio – > GoldenRatio,
    PlotStyle – > {{RGBColor [1, 0, 0 ], Thickness [0 .01]}}]
```



Solve for the cut score for group A given that *m* total applicants are to be selected and that the desired representation of group A is p.

$C_A[p\_, m\_] = C_A /. \text{Solve}[n_A (1 - F_A[C_A]) == p\, m, C_A][[1]]$

Solve: : ifun : Inverse functions are being used by Solve, so some solutions may not be found.

$$\mu_A + \sqrt{2}\,\text{InverseErf}\left[0, 1 - \frac{2mp}{n_A}\right]\sigma_A$$

Solve for the cut score for group B given that m total applicants are to be selected and that the representation of group A is p.

$C_B[p\_, m\_] = C_B /. \text{Solve}[n_B (1 - F_B[C_B]) == (1 - p)\, m, C_B][[1]]$

Solve: : ifun : Inverse functions are being used by Solve, so some solutions may not be found.

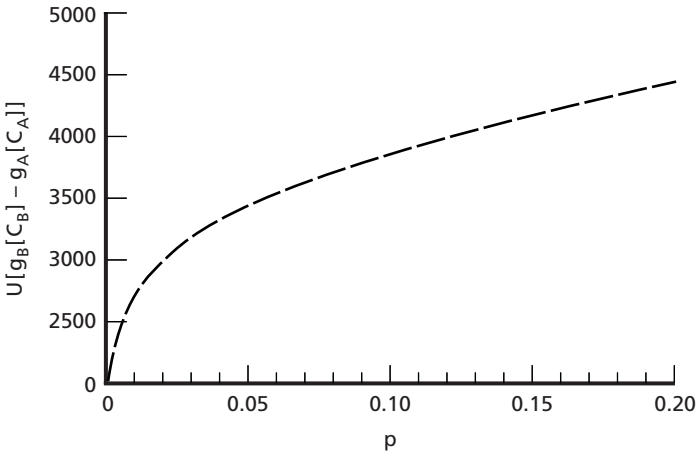
$$\mu_B + \sqrt{2}\,\text{InverseErf}\left[0, \frac{-2m + 2mp + n_B}{n_B}\right]\sigma_B$$

Define a utility function for differences in marginal performance.

```
U[o_] := 200o
```

Plot the cost at the margin in terms of the difference in performance versus the representation of group A.

```
Plot[Evaluate[U[gB[CB[p, 20]] - gA[CA[p, 20]]] /.
  Parameters], {p, 0.00, 0.20}, PlotPoints → 40,
    PlotRange → {2000, 5000}, AspectRatio → GoldenRatio,
      AxesLabel → {p, "U[gB[CB] - gA[CA]]"}, PlotStyle →
        {{RGBColor[1, 0, 0], Thickness[0.02]}}]
```


5000
4500
4000
3500
3000
2500
0
U[g_B[C_B] – g_A[C_A]]
0
0.05
0.10
0.15
0.20
p