INSIDE

Kaufmann and Kraay on Governance Indicators

Devarajan and Johnson
Comment on Kaufmann/Kraay

Easterly, Irwin, and Servén on
Public Investment and Fiscal Stability

Gill, Ozer, and Tatucu on Social Security Reform

Cadot and de Melo on Reforming Rules of Origin

ISSN 0257-3032
www.wbro.oxfordjournals.org
The World Bank Research Observer is intended for anyone who has a professional interest in development. Observer articles are written to be accessible to nonspecialist readers; contributors examine key issues in development economics, survey the literature and the latest World Bank research, and debate issues of development policy. Articles are reviewed by an editorial board drawn from across the Bank and the international community of economists. Inconsistency with Bank policy is not grounds for rejection.

The journal welcomes editorial comments and responses, which will be considered for publication to the extent that space permits. On occasion the Observer considers unsolicited contributions. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages to the Editor. Please direct all editorial correspondence to the Editor, The World Bank Research Observer, 1818 H Street, NW, Washington, DC 20433, USA.

The views and interpretations expressed in this journal are those of the authors and do not necessarily represent the views and policies of the World Bank or of its Executive Directors or the countries they represent. The World Bank does not guarantee the accuracy of data included in this publication and accepts no responsibility whatsoever for any consequences of their use. When maps are used, the boundaries, denominations, and other information do not imply on the part of the World Bank Group any judgment on the legal status of any territory or the endorsement or acceptance of such boundaries.

Governance Indicators: Where Are We, Where Should We Be Going?
Daniel Kaufmann and Aart Kraay

Two Comments on "Governance Indicators: Where Are We, Where Should We Be Going?" by Daniel Kaufmann and Aart Kraay
Shantayanan Devarajan and Simon Johnson

Walking up the Down Escalator: Public Investment and Fiscal Stability
William Easterly, Timothy Irwin, and Luis Servén

What Can Countries in Other Regions Learn from Social Security Reform in Latin America?
Indermit S. Gill, Ceren Ozer, and Radu Tatucu

Why OECD Countries Should Reform Rules of Origin
Olivier Cadot and Jaime de Melo
Governance Indicators: Where Are We, Where Should We Be Going?

Daniel Kaufmann • Aart Kraay

Progress in measuring governance is assessed using a simple framework that distinguishes between indicators that measure formal rules and indicators that measure the practical application or outcomes of these rules. The analysis calls attention to the strengths and weaknesses of both types of indicators as well as the complementarities between them. It distinguishes between the views of experts and the results of surveys and assesses the merits of aggregate as opposed to individual governance indicators. Some simple principles are identified to guide the use and refinement of existing governance indicators and the development of future indicators. These include transparently disclosing and accounting for the margins of error in all indicators, drawing from a diversity of indicators and exploiting complementarities among them, submitting all indicators to rigorous public and academic scrutiny, and being realistic in expectations of future indicators. JEL codes: H1, O17

Not everything that can be counted counts, and not everything that counts can be counted. —Albert Einstein

Most scholars, policymakers, aid donors, and aid recipients recognize that good governance is a fundamental ingredient of sustained economic development. This growing understanding, initially informed by a very limited set of empirical measures of governance, has spurred intense interest in developing more refined, nuanced, and policy-relevant indicators of governance. This article reviews progress in measuring governance, emphasizing empirical measures explicitly designed to be comparable across countries and in most cases over time. The goal is to provide a structure for thinking about the strengths and weaknesses of different types of governance indicators that can inform both the use of existing indicators and ongoing efforts to improve them and develop new ones.¹

The first section of this article reviews definitions of governance. Although there are many broad definitions of governance, the degree of definitional disagreement
can easily be overstated. Most definitions appropriately emphasize the importance of a capable state that is accountable to citizens and operating under the rule of law. Broad principles of governance along these lines are naturally not amenable to direct observation and thus to direct measurement. As Albert Einstein noted, "Not everything that counts can be counted." Many different types of data provide information on the extent to which these principles of governance are observed across countries. An important corollary is that any particular indicator of governance can usefully be interpreted as an imperfect proxy for some unobserved broad dimension of governance. This interpretation emphasizes throughout this review a recurrent theme that there is measurement error in all governance indicators, which should be explicitly considered when using these kinds of data to draw conclusions about cross-country differences or trends in governance over time.

The second section addresses what is measured. The discussion highlights the distinction between indicators that measure specific rules "on the books" and indicators that measure particular governance outcomes "on the ground." Rules on the books codify details of the constitutional, legal, or regulatory environment: the existence or absence of specific agencies, such as anticorruption commissions or independent auditors; and so forth—components intended to provide the key de jure foundations of governance. On-the-ground measures assess de facto governance outcomes that result from the application of these rules (Do firms find the regulatory environment cumbersome? Do households believe the police are corrupt?). An important message in this section concerns the shared limitations of indicators of both rules and outcomes: Outcome-based indicators of governance can be difficult to link back to specific policy interventions, and the links from easy-to-measure de jure indicators of rules to governance outcomes of interest are not yet well understood and in some cases appear tenuous at best. They remind us of the need to respect Einstein's dictum that "not everything that can be counted counts."

The third section examines whose views should be relied on. Indicators based on the views of various types of experts are distinguished from survey-based indicators that capture the views of large samples of firms and individuals. A category of aggregate indicators that combine, organize, provide structure, and summarize information from these different types of respondents is examined. The fourth section examines the rationale for such aggregate indicators, and their strengths and weaknesses.

The set of indicators discussed in this survey is intended to provide leading examples of major governance indicators rather than an exhaustive stocktaking of existing indicators in this taxonomy. A feature of efforts to measure governance is the preponderance of indicators focused on measuring de facto governance outcomes and the paucity of measures of de jure rules. Almost by necessity, de jure rules-based indicators of governance reflect the views or judgments of experts. In
contrast, the much larger body of *de facto* indicators captures the views of both experts and survey respondents.

The article concludes with a discussion of the way forward in measuring governance in a manner that can be useful to policymakers. The emphasis is on the importance of consumers and producers of governance indicators clearly recognizing and disclosing the pervasive measurement error in any type of governance indicators. *This section also notes* the importance of moving away from oft-heard false dichotomies, such as "subjective" or "objective" indicators or aggregate or disaggregated ones. For good reason, virtually all measures of governance involve a degree of subjective judgment, and different levels of aggregation are appropriate for different types of analysis. In any case, the choice is not either one or the other, as most aggregate indicators can readily be unbundled into their constituent components.

**What Does Governance Mean?**

The concept of governance is not a new one. Early discussions go back to at least 400 BCE, to the *Arthashastra*, a treatise on governance attributed to Kautilya, thought to be the chief minister to the king of India. Kautilya presents key pillars of the "art of governance," emphasizing justice, ethics, and anti-autocratic tendencies. He identifies the duty of the king to protect the wealth of the state and its subjects and to enhance, maintain, and safeguard this wealth as well as the interests of the kingdom's subjects.

Despite the long provenance of the concept, no strong consensus has formed around a single definition of governance or institutional quality. For this reason, throughout this article the terms *governance*, *institutions*, and *institutional quality* are used interchangeably, if somewhat imprecisely. Researchers and organizations have produced a wide array of definitions. Some definitions are so broad that they cover almost anything (such as the definition "rules, enforcement mechanisms, and organizations" offered in the World Bank's *World Development Report 2002: Building Institutions for Markets*). Others, like the definition suggested by North (2000), are not only broad but risk making the links from good governance to development almost tautological: "How do we account for poverty in the midst of plenty? . . . We must create incentives for people to invest in more efficient technology, increase their skills, and organize efficient markets . . . Such incentives are embodied in institutions."

Some of the governance indicators surveyed capture a wide range of development outcomes. While it is difficult to draw a line between governance and the ultimate development outcomes of interest, it is useful at both the definitional and measurement stages to emphasize concepts of governance that are at least somewhat removed from development outcomes themselves. An early and narrower
definition of public sector governance proposed by the World Bank is that “governance is the manner in which power is exercised in the management of a country’s economic and social resources for development” (World Bank 1992, p. 1). This definition remains almost unchanged in the Bank’s 2007 governance and anticorruption strategy, with governance defined as “the manner in which public officials and institutions acquire and exercise the authority to shape public policy and provide public goods and services” (World Bank 2007, p. 1).

Kaufmann, Kraay, and Zoido-Lobatón (1999a, p. 1) define governance as “the traditions and institutions by which authority in a country is exercised. This includes the process by which governments are selected, monitored and replaced; the capacity of the government to effectively formulate and implement sound policies; and the respect of citizens and the state for the institutions that govern economic and social interactions among them.”

Although the number of definitions of governance is large, there is some consensus. Most definitions agree on the importance of a capable state operating under the rule of law. Interestingly, comparing the last three definitions cited above, the one substantive difference has to do with the explicit degree of emphasis on the role of democratic accountability of governments to their citizens. Even these narrower definitions remain sufficiently broad that there is scope for a wide diversity of empirical measures of various dimensions of good governance.

The gravity of the issues dealt with in these definitions of governance suggests that measurement is important. In recent years there has been debate over whether such broad notions of governance can be usefully measured. Many indicators can shed light on various dimensions of governance. However, given the breadth of the concepts, and in many cases their inherent unobservability, no single indicator or combination of indicators can provide a completely reliable measure of any of these dimensions of governance. Rather, it is useful to think of the various specific indicators discussed below as all providing imperfect signals of fundamentally unobservable concepts of governance. This interpretation emphasizes the importance of taking into account as explicitly as possible the inevitable resulting measurement error in all indicators of governance when analyzing and interpreting any such measure. As shown below, however, the fact that such margins of error are finite and still allow for meaningful country comparisons across space and time suggests that measuring governance is both feasible and informative.

**Governance Rules or Governance Outcomes?**

This section examines both the rules-based and outcome-based indicators of governance. A rules-based indicator of corruption might measure whether countries have legislation prohibiting corruption or have an anticorruption agency.
An outcome-based measure could assess whether the laws are enforced or the anticorruption agency is undermined by political interference. The views of firms, individuals, nongovernmental organizations (NGOs), or commercial risk-rating agencies could also be solicited regarding the prevalence of corruption in the public sector. To measure public sector accountability, one could observe the rules regarding the presence of formal elections, financial disclosure requirements for public servants, and the like. One could also assess the extent to which these rules operate in practice by surveying respondents regarding the functioning of the institutions of democratic accountability.

Because a clear line does not always distinguish the two types of indicators, it is more useful to think of ordering different indicators along a continuum, with one end corresponding to rules and the other to ultimate governance outcomes of interest. Because both types of indicators have their own strengths and weaknesses, all indicators should be thought of as imperfect, but complementary proxies for the aspects of governance they purport to measure.

**Rules-Based Indicators of Governance**

Several rules-based indicators are used to assess governance (tables 1 and 2). They include the Doing Business project of the World Bank, which reports detailed information on the legal and regulatory environment in a large set of countries; the Database of Political Institutions, constructed by World Bank researchers, and the POLITY-IV database of the University of Maryland, both of which report detailed factual information on features of countries' political systems; and the Global Integrity Index (GII), which provides detailed information on the legal framework governing public sector accountability and transparency in a sample of 41 countries, most of them developing economies.

At first glance, one of the main virtues of indicators of rules is their clarity. It is straightforward to ascertain whether a country has a presidential or a parliamentary system of government or whether a country has a legally independent anticorruption commission. In principle, it is also straightforward to document details of the legal and regulatory environment, such as how many legal steps are required to register a business or fire a worker. This clarity also implies that it is straightforward to measure progress on such indicators. Has an anticorruption commission been established? Have business entry regulations been streamlined? Has a legal requirement for disclosure of budget documents been passed? This clarity has made such indicators very appealing to aid donors interested in linking aid with performance indicators and in monitoring progress on such indicators. Set against these advantages are three main drawbacks.

*They are less “objective” than they appear.* It is easy to overstate the clarity and objectivity of rules-based measures of governance. In practice, a good deal of
subjective judgment is involved in codifying all but the most basic and obvious features of a country’s constitutional, legal, and regulatory environments. (It is no accident that the views of lawyers, on which many of these indicators are based, are commonly referred to as opinions.) In Kenya in 2007, for example, a constitutional right to access to information faced being undermined or offset entirely by an official secrecy act and by pending approval and implementation of the Freedom of Information Act. In this case, codifying even the legal right to access to information requires careful judgment as to the net effect of potentially conflicting laws. Of course, this drawback of ambiguity is not unique to rules-based measures of governance: interpreting outcome-based indicators of governance can also involve ambiguity, as discussed below. There has been less recognition, however, of the extent to which rules-based indicators also reflect subjective judgment.

The links between indicators and outcomes are complex, possibly subject to long lags, and often not well understood. These problems complicate the interpretation of rules-based indicators. In the case of rules-based measures, some of the most basic features of countries’ constitutional arrangements have little normative content on their own: such indicators are for the most part descriptive. It makes

Table 1. Sources and Types of Information Used in Governance Indicators

<table>
<thead>
<tr>
<th>Source of Information</th>
<th>Rules-based</th>
<th>Outcomes-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Broad</td>
<td>Specific</td>
</tr>
<tr>
<td>Experts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawyers</td>
<td>DB</td>
<td>DRI, EIU, PRS</td>
</tr>
<tr>
<td>Commercial risk-rating agencies</td>
<td>GII</td>
<td>HER, RSE, CIR, FRH</td>
</tr>
<tr>
<td>Nongovernmental organizations</td>
<td>CPIA</td>
<td>DRI, EIU, PRS</td>
</tr>
<tr>
<td>Governments and multilaterals</td>
<td>DPI, PIV</td>
<td>DRI, EIU, PRS</td>
</tr>
<tr>
<td>Academics</td>
<td>DPI, PIV</td>
<td>DRI, EIU, PRS</td>
</tr>
<tr>
<td>Survey respondents</td>
<td>DPI, PIV</td>
<td>DRI, EIU, PRS</td>
</tr>
<tr>
<td>Firms</td>
<td>ICA, GCS, WCY</td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>AFR, LBO, GWP</td>
<td></td>
</tr>
<tr>
<td>Aggregate indicators combining experts and survey respondents</td>
<td>TI, WGI, MOI</td>
<td></td>
</tr>
</tbody>
</table>

Note: AFR is Afrobarometer, CIR is Cingranelli-Richards Human Rights Dataset, CPIA is Country Policy and Institutional Assessment, DB is Doing Business, DPI is Database of Political Institutions, DRI is Global Insight DRI, EIU is Economist Intelligence Unit, FRH is Freedom House, GCS is Global Competitiveness Survey, GII is Global Integrity Index, GWP is Gallup World Poll, HER is Heritage Foundation, ICA is Investment Climate Assessment, LBO is Latinobarométre, MOI is Ibrahim Index of African Governance, OBI is Open Budget Index, PEFA is Public Expenditure and Financial Accountability, PIV is Polity IV, PRS is Political Risk Services, RSF is Reporters Without Borders, TI is Transparency International, WCY is World Competitiveness Yearbook, and WGI is Worldwide Governance Indicators.

Source: Authors’ compilation based on data from sources listed in table 2.

Table 2. Country Coverage and Frequency of Governance Surveys

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of countries covered</th>
<th>Frequency of survey</th>
<th>Web site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afrobarometer</td>
<td>18</td>
<td>Triennial</td>
<td><a href="http://www.afrobarometer.org">www.afrobarometer.org</a></td>
</tr>
<tr>
<td>Cingranelli-Richards Human Rights Dataset</td>
<td>192</td>
<td>Annual</td>
<td><a href="http://www.humanrightsdata.com">www.humanrightsdata.com</a></td>
</tr>
<tr>
<td>Doing Business</td>
<td>175</td>
<td>Annual</td>
<td><a href="http://www.doingbusiness.org">www.doingbusiness.org</a></td>
</tr>
<tr>
<td>Database of Political Institutions</td>
<td>178</td>
<td>Annual</td>
<td><a href="http://econ.worldbank.org">http://econ.worldbank.org</a></td>
</tr>
<tr>
<td>Global Insight DRI</td>
<td>117</td>
<td>Quarterly</td>
<td><a href="http://www.globalinsight.com">www.globalinsight.com</a></td>
</tr>
<tr>
<td>Economist Intelligence Unit</td>
<td>120</td>
<td>Quarterly</td>
<td><a href="http://www.eiu.com">www.eiu.com</a></td>
</tr>
<tr>
<td>Freedom House</td>
<td>192</td>
<td>Annual</td>
<td><a href="http://www.freedomhouse.org">www.freedomhouse.org</a></td>
</tr>
<tr>
<td>Global Competitiveness Survey</td>
<td>117</td>
<td>Annual</td>
<td><a href="http://www.weforum.org">www.weforum.org</a></td>
</tr>
<tr>
<td>Global Integrity Index</td>
<td>41</td>
<td>Triennial</td>
<td><a href="http://www.globalintegrity.org">www.globalintegrity.org</a></td>
</tr>
<tr>
<td>Gallup World Poll</td>
<td>131</td>
<td>Annual</td>
<td><a href="http://www.gallupworldpoll.com">www.gallupworldpoll.com</a></td>
</tr>
<tr>
<td>Heritage Foundation</td>
<td>161</td>
<td>Annual</td>
<td><a href="http://www.heritage.org">www.heritage.org</a></td>
</tr>
<tr>
<td>Investment Climate Assessment</td>
<td>94</td>
<td>Irregular</td>
<td><a href="http://www.investmentclimate.org">www.investmentclimate.org</a></td>
</tr>
<tr>
<td>Latinobarometro</td>
<td>17</td>
<td>Annual</td>
<td><a href="http://www.latinobarometro">www.latinobarometro</a></td>
</tr>
<tr>
<td>Ibrahim Index of African Governance</td>
<td>48</td>
<td>Triennial</td>
<td><a href="http://www.molibrahimfoundation.org">www.molibrahimfoundation.org</a></td>
</tr>
<tr>
<td>Open Budget Index</td>
<td>59</td>
<td>Annual</td>
<td><a href="http://www.openbudgetindex.org">www.openbudgetindex.org</a></td>
</tr>
<tr>
<td>Polity IV</td>
<td>161</td>
<td>Annual</td>
<td><a href="http://www.cidcm.umd.edu/polity/">www.cidcm.umd.edu/polity/</a></td>
</tr>
<tr>
<td>Political Risk Services</td>
<td>140</td>
<td>Monthly</td>
<td><a href="http://www.prsgroup.com">www.prsgroup.com</a></td>
</tr>
<tr>
<td>Public Expenditure and Financial Accountability</td>
<td>42</td>
<td>Irregular</td>
<td><a href="http://www.pefa.org">www.pefa.org</a></td>
</tr>
<tr>
<td>Reporters without Borders</td>
<td>165</td>
<td>Annual</td>
<td><a href="http://www.rsf.org">www.rsf.org</a></td>
</tr>
<tr>
<td>World Competitiveness Yearbook</td>
<td>47</td>
<td>Annual</td>
<td><a href="http://www.imd.ch">www.imd.ch</a></td>
</tr>
</tbody>
</table>

Source: Authors' compilation.

little sense, for example, to presuppose that presidential (as opposed to parliamentary) systems or majoritarian (as opposed to proportional) representation in voting arrangements are intrinsically good or bad. Interest in such variables as indicators of governance rests on the case that they may matter for outcomes, often in complex ways. In their influential book, Persson, Torsten, and Tabellini (2005) document how these features of constitutional rules influence the political process and ultimately outcomes such as the level, composition, and cyclicity of public spending (Acemoglu 2006) challenges the robustness of these findings). In such cases, the usefulness of rules-based indicators as measures of governance depends crucially on how strong the empirical links are between such rules and the ultimate outcomes of interest.

Perhaps the more common is the less extreme case in which rules-based indicators of governance have normative content on their own. but the relative
importance of different rules for outcomes of interest is unclear. The GII, for example, provides information on the existence of dozens of rules, ranging from the legal right to freedom of speech to the existence of an independent ombudsman to the presence of legislation prohibiting the offering or acceptance of bribes. The Open Budget Index (OBI) provides detailed information on the budget processes, including the types of information provided in budget documents, public access to budget documents, and the interaction between executive and legislative branches in the budget process. Many of these indicators arguably have normative value on their own: having public access to budget documents is desirable and having streamlined business registration procedures is better than not having them.

This profusion of detail in rules-based indicators leads to two related difficulties in using them to design and monitor governance reforms. The first is that as a result of absence of good information on the links between changes in specific rules or procedures and outcomes of interest, it is difficult to know which rules should be reformed and in what order. Will establishing an anticorruption commission or passing legislation outlawing bribery have any impact on reducing corruption? If so, which is more important? Should, instead, more efforts be put into ensuring that existing laws and regulations are implemented or that there is greater transparency, access to information, or media freedom? How soon should one expect to see the impacts of these interventions? Given that governments typically operate with limited political capital to implement reforms, these trade-offs and lags are important.

The second difficulty in designing or monitoring reforms arises when aid donors or governments set performance indicators for governance reforms. Performance indicators based on changing specific rules, such as the passage of a particular piece of legislation or a reform of a specific budget procedure, can be very attractive because of their clarity: it is straightforward to verify whether the specified policy action has been taken. Yet “actionable” indicators are not necessarily also “action worthy,” in the sense of having a significant impact on the outcomes of interest. Moreover, excessive emphasis on registering improvements on rules-based indicators of governance leads to risks of “teaching to the test” or, worse “reform illusion,” in which specific rules or procedures are changed in isolation with the sole purpose of showing progress on the specific indicators used by aid donors.

Major gaps exist between statutory rules on the books and their implementation on the ground. To take an extreme example, in all 41 countries covered by the 2006 GII, accepting a bribe is codified as illegal, and all but three countries (Brazil, Lebanon, and Liberia) have anticorruption commissions or similar agencies. Yet there is enormous variation in perceptions-based measures of corruption across these countries. The 41 countries covered by the GII include the Democratic
Republic of Congo, which ranks 200 out of 207 countries on the 2006 Worldwide Governance Indicators (WGI) control of corruption indicator, and the United States, which ranks 23.

Another example of the gap between rules and implementation (documented in more detail in Kaufmann, Kraay, and Mastruzzi 2005) compares the statutory ease of establishing a business with a survey-based measure of firms’ perceptions of the ease of starting a business across a large sample of countries. In industrial countries, where de jure rules are often implemented as intended, the two measures correspond quite closely. In contrast, in developing economies, where there are often gaps between de jure rules and their de facto implementation, the correlation between the two is very weak; the de jure codification of the rules and regulations required to start a business is not a good predictor of the actual constraints reported by firms. Unsurprisingly, much of the difference between the de jure and de facto measures could be statistically explained by de facto measures of corruption, which subverts the fair application of rules on the books.

The three drawbacks—the inevitable role of judgment even in “objective” indicators, the complexity and lack of knowledge regarding the links from rules to outcomes of interest, and the gap between rules on the books and their implementation on the ground—suggest that although rules-based governance indicators provide valuable information, they are insufficient on their own for measuring governance. Rules-based measures need to be complemented by and used in conjunction with outcome-based indicators of governance.

**Outcome-Based Governance Indicators**

Most indicators of governance are outcome based, and several rules-based indicators of governance also provide complementary outcome-based measures. The GII, for example, pairs indicators of the existence of various rules and procedures with indicators of their effectiveness in practice. The Database of Political Institutions measures not only such constitutional rules as the presence of a parliamentary system, but also outcomes of the electoral process, such as the extent to which one party controls different branches of government and the fraction of votes received by the president. The Polity-IV database records a number of outcomes, including the effective constraints on the power of the executive.

The remaining outcome-based indicators range from the highly specific to the quite general. The OBI reports data on more than 100 indicators of the budget process, ranging from whether budget documentation contains details of assumptions underlying macroeconomic forecasts to documentation of budget outcomes relative to budget plans. Other less specific sources include the Public Expenditure and Financial Accountability indicators, constructed by aid donors with inputs of recipient countries, and several large cross-country surveys of firms—including
the Investment Climate Assessments of the World Bank, the Executive Opinion Survey of the World Economic Forum, and the World Competitiveness Yearbook of the Institute for Management Development—that ask firms detailed questions about their interactions with the state.

Examples of more general assessments of broad areas of governance include ratings provided by several commercial sources, including Political Risk Services, the Economist Intelligence Unit, and Global Insight–DRI. Political Risk Services rates 10 areas that can be identified with governance, such as "democratic accountability," "government stability," "law and order," and "corruption." Large cross-country surveys of individuals such as the Afrobarometer and Latinobarómetro surveys and the Gallup World Poll ask general questions, such as "Is corruption widespread throughout the government in this country?"

The main advantage of outcome-based indicators is that they capture the views of relevant stakeholders, who take actions based on these views. Governments, analysts, researchers, and decisionmakers should, and often do, care about public views on the prevalence of corruption, the fairness of elections, the quality of service delivery, and many other governance outcomes. Outcome-based governance indicators provide direct information on the de facto outcome of how de jure rules are implemented.

Outcome-based measures also have some significant limitations. Such measures, particularly where they are general, can be difficult to link back to specific policy interventions that might influence governance outcomes. This is the mirror image of the problem discussed above: Rules-based indicators of governance can also be difficult to relate to outcomes of interest. A related difficulty is that outcome-based governance indicators may be too close to ultimate development outcomes of interest. To take an extreme example, the Ibrahim Index of African Governance includes a number of ultimate development outcomes, such as per capita GDP (gross domestic product), growth of GDP, inflation, infant mortality, and inequality. While such development outcomes are surely worth monitoring, including them in an index of governance risks making the links from governance to development tautological.

Another difficulty has to do with interpreting the units in which outcomes are measured. Rules-based indicators have the virtue of clarity: either a particular rule exists or it does not. Outcome-based indicators by contrast are often measured on somewhat arbitrary scales. For example, a survey question might ask respondents to rate the quality of public services on a five-point scale, with the distinction between different scores left unclear and up to the respondent. In contrast, the usefulness of outcome-based indicators is greatly enhanced by the extent to which the criteria for differing scores are clearly documented. The World Bank's Country Performance and Institutional Assessment (CPIA) and the Freedom House indicators are good examples of outcome-based indicators based
on expert assessments that provide documentation of the criteria used to assign specific scores on the indicators they compile. In the case of surveys, questions can be designed to ensure that responses are easier to interpret: rather than asking respondents whether they think "corruption is widespread," respondents can be asked whether they have been solicited for a bribe in the past month.

An example illustrates some of the main advantages and disadvantages of the two types of measures. Figure 1 compares alternative indicators of democratic accountability, a key dimension of governance. The horizontal axis measures a very broad outcome-based indicator, taken from the 2005 Voice of the People survey, a large cross-country household survey (www.voice-of-the-people.net). It asks households to indicate whether they think elections in their country are free and fair. The vertical axis reports two indicators of the quality of electoral institutions, taken from Global Integrity. The points labeled "de jure" are based on a factual assessment of the existence of a number of specific institutions related to

![Figure 1. De facto and de jure Indicators of Elections](image)

**De jure**

\[ y = -3.14x + 85.69 \]

\[ R^2 = 0.00 \]

**De facto**

\[ y = 23.09x + 55.30 \]

\[ R^2 = 0.19 \]

**Voice of the People Household Survey:** Are elections free and fair?

**Note:** ARG is Argentina, ARM is Armenia, AZE is Azerbaijan, BEN is Benin, BRA is Brazil, BGR is Bulgaria, ZAR is Democratic Republic of Congo, EGY is Egypt, ETH is Ethiopia, GEO is Georgia, GHA is Ghana, GTM is Guatemala, IND is India, IDN is Indonesia, ISR is Israel, KEN is Kenya, KGZ is Kyrgyz Republic, LBN is Lebanon, LBR is Liberia, MEX is Mexico, MNP is Montenegro, MOZ is Mozambique, NPR is Nepal, NIC is Nicaragua, NGA is Nigeria, PAK is Pakistan, PHL is Philippines, ROM is Romania, RUS is Russia, SEN is Senegal, YUG is Serbia, SLE is Sierra Leone, ZAF is South Africa, SDN is Sudan, TJK is Tajikistan, TZA is Tanzania, UGA is Uganda, USA is United States, VNM is Vietnam, YEM is Yemen, and ZWE is Zimbabwe.

**Source:** Authors' analysis based on data described in the text.
elections, such as the existence of a legal right to universal suffrage and the existence of an election monitoring agency. The points labeled "de facto" capture the assessment of Global Integrity’s experts as to the effectiveness of these institutions.  

Several messages emerge from this figure. First, in some cases rules-based measures of governance show remarkably little variation across countries, with all countries receiving scores close to 100, indicating perfect scores on the de jure basis of this important aspect of governance. As of 2005, for example, every country surveyed by Global Integrity promised the legal right to vote, and a statutorily independent election-monitoring agency existed in all but three countries (Lebanon, Montenegro, and Mozambique). Second, the links between a specific objective indicator of rules and the broad outcome of interest (citizens’ satisfaction with elections) is at best very weak, with a correlation between the two measures that is in fact slightly negative. Third, outcome-based indicators explicitly focusing on the de facto implementation of rules can be useful. A noteworthy feature of Global Integrity is its pairing of indicators of specific rules with assessments of their functioning in practice. The correlation of the de facto measure with the broad outcome measure of interest taken from the Voice of the People survey is much stronger (0.46) than the correlation with the de jure measure. The correlation is far from perfect, however, indicating the importance of relying on a variety of indicators when assessing governance in a country.

Whose Views Should We Rely On?

A variety of governance assessments are produced by experts on behalf of commercial risk-rating agencies and NGOs. The GII and the OBI, for example, rely on locally recruited experts in each country to complete their detailed questionnaires about governance, subject to peer review. Commercial organizations such as the Economist Intelligence Unit rely on a network of local correspondents in a large set of countries to provide information underlying the ratings they produce. Other advocacy organizations, such as Amnesty International, Freedom House, and Reporters without Borders, also rely on networks of respondents for the information underlying their assessments.

Governments and multilateral organizations are also major producers of expert assessments. Some of the most notable include the Country Policy and Institutional Assessments, produced by the World Bank, the African Development Bank, and the Asian Development Bank. Each of these assessments is based on the responses of each institution’s country economists to a detailed questionnaire, responses that are then reviewed for consistency and comparability across countries. The Public
Expenditure and Financial Accountability indicators mentioned earlier are also based on experts’ views.

Several large cross-country surveys of firms and individuals contain questions on governance. These include the Investment Climate Assessment and the Business Environment and Enterprise Performance Surveys conducted by the World Bank; the Executive Opinion Survey of the World Economic Forum; the World Competitiveness Yearbook; Voice of the People; and the Gallup World Poll.

**Expert Assessments**

Expert assessments have several major advantages, which account for their preponderance among various types of governance indicators. One is cost: it is much less expensive to ask a selection of country economists at the World Bank to provide responses to a questionnaire on governance as part of the CPIA process than to carry out representative surveys of firms or households in a hundred or more countries. The second advantage is that expert assessments can more readily be tailored for cross-country comparability: Many of the organizations listed in table 2 have elaborate benchmarking systems to ensure that scores are comparable across countries. Finally, for certain aspects of governance experts are the natural respondents for the type of information being sought. (Consider, for example, the OBIE’s detailed questionnaire on national budget processes, the particulars of which are not the sort of common knowledge that survey data can easily collect.)

Expert assessments nevertheless have several important limitations. A basic one is that, like survey respondents, different experts may have different views about similar aspects of governance. While this is not surprising, it suggests that users of governance indicators should be cautious about relying too heavily on any one set of expert assessments. These differences are evident in comparing the CPIA ratings of the World Bank and the African Development Bank, which in recent years harmonized their procedures for constructing CPIA ratings. An identical questionnaire covering 16 dimensions of policy and institutional performance is completed by two very similar sets of expert respondents—country economists with in-depth experience working on behalf of these two organizations in the countries they are assessing. Despite the homogeneity of the respondents and the very similar rating criteria, there are nontrivial differences between the two organizations’ assessments on the 16 components of the CPIA (table 3). For example, the 0.67 correlation between the two assessments on the question on transparency, accountability, and corruption in the public sector is far from perfect, suggesting that it is prudent to base assessments of governance for policy purposes on the views of a variety of expert assessments.\(^4\)
Table 3. Correlation Among Alternative Indicators of Corruption

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Expert assessments</th>
<th>Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>World Bank CPIA</td>
<td>African Development Bank CPIA</td>
</tr>
<tr>
<td>World Bank CPIA</td>
<td>1.00</td>
<td>0.67</td>
</tr>
<tr>
<td>African Development Bank CPIA</td>
<td>1.00</td>
<td>0.49</td>
</tr>
<tr>
<td>Global Integrity</td>
<td>1.00</td>
<td>0.34</td>
</tr>
<tr>
<td>World Markets Online</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>World Economic Forum Executive Opinion Survey</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Gallup World Poll</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Source: Authors’ analysis based on data described in the text.*

The second criticism that the country ratings assigned by different groups of experts are too highly correlated is just the opposite. Suppose that one set of experts comes up with an assessment of governance for a set of countries based on its own independent research and the second set of experts simply reproduces the assessments of the first. In this case, the high correlation of two expert assessments cannot be interpreted as evidence of their accuracy. Rather, it would reflect the fact that the two sources make correlated errors in measuring governance.7

Nevertheless, even if the errors made by two data sources are highly, but not perfectly correlated, there will be benefits to relying on both data sources. The important empirical question is whether this hypothetical correlation of errors across sources is large or not. Empirically identifying correlations in errors across sources is difficult. Simply observing whether the assessments provided in the two data sources are highly correlated is not enough, as the high correlation can reflect the fact that both sources are either measuring governance accurately or making correlated measurement errors.

To make progress, one needs to make identifying assumptions. Kaufmann, Kraay, and Mastruzzi (2006) detail two sets of assumptions that allow potential sources of correlation in the errors to be disentangled. One is that surveys of firms or individuals are less likely to make errors that are correlated with other data sources than, for example, assessments by commercial risk-rating agencies. If this is the case, however, one would expect that the assessments of commercial risk-rating agencies would be very highly correlated with one another, but less so with surveys. This turns out not to be the case. The average correlation of the five
major commercial risk-rating agencies for corruption in 2002–05 was 0.80. The correlation of each of these assessments with a large cross-country survey of firms was slightly higher (0.81), in contrast with what one would expect if the rating agencies had correlated errors. Conducting this exercise across all six aggregate governance indicators reveals at most modest evidence of error correlation. While this is unlikely to be the final word on this important question, it is a useful step forward to propose and implement tests of error correlation based on explicit identifying assumptions.

The third criticism is that expert assessments are subject to various biases. Some researchers claim that many of these sources are biased toward the views of the business community, which may have very different views of what constitutes good governance than do other types of respondents. In short, goes the critique, businesspeople like low taxes and less regulation, while the public good demands reasonable taxation and appropriate regulation. This critique does not seem particularly compelling. If it were true, the responses of commercial risk-rating agencies, which serve mostly business clients, or the views of firms themselves to questions about governance, should not be highly correlated with ratings provided by respondents who are more likely to sympathize with the common good, such as individuals, NGOs, or public sector organizations. Yet, in most cases, these correlations are strong (Kaufmann, Kraay, and Mastruzzi 2007b). Cross-country surveys of firms and of individuals, such as the World Economic Forum’s Executive Opinion Survey and the Gallup World Poll, yield similar corruption rankings, with the two surveys correlated at 0.7 (table 3).

Another potential source of bias in expert assessments, particularly those produced by NGOs, is that they are colored by the ideological orientation of the ratings organization. Kaufmann, Kraay, and Mastruzzi (2004) find that the assessments of think tanks and firm surveys are not systematically correlated with the political orientation of a country’s government, casting doubt on this possible source of bias. A potentially greater problem of bias is at the country respondent level. For example, the views of pro-government and antigovernment “experts” might be very different, affecting both levels and trends over time. This risk is perhaps greatest for the sources that rely on local experts, such as the GII. This risk is also much more difficult to test for systematically, as the biases may affect individual country scores without introducing systematic biases into the source as a whole. Nevertheless, careful comparisons of many different data sources can often turn up anomalies in a single source that require more careful scrutiny.

**Surveys of Firms and Individuals**

Governance indicators derived from surveys of firms and individuals have the fundamental advantage that they elicit the views of the ultimate beneficiaries of good
governance, citizens and firms in a country. The views of these stakeholders matter because they are likely to act on those views. If firms or individuals believe that the courts and the police are corrupt, they are unlikely to try to use their services (Hellman and Kaufmann 2004). Individuals are less likely to vote or to hold their elected leaders accountable if they think that elections are not free and fair.

Another advantage of governance indicators based on surveys of domestic firms and individuals is greater domestic political credibility. Governments often dismiss external expert assessments of governance as uninformed pontification by outsiders. It is much harder for them to dismiss the views of their own citizens or of firms operating in their country. Survey-based data on governance can therefore be particularly useful in galvanizing the politics of governance reforms. The experience of many countries implementing their own in-depth Governance and Anti-Corruption diagnostics (assisted by the World Bank Institute and other agencies and implemented with institutions in the requesting country), based on in-country surveys of enterprises, users of services, and public officials, supports this point: the views expressed by thousands of domestic stakeholders provide powerful input for action to reformist policymakers and civil society groups.

Set against these important advantages of surveys are a number of disadvantages. First, there is the usual array of potential problems with any type of survey data, ranging from issues of sampling design to issues of nonresponse bias. Expert assessments, which are based on the views of a very small number of respondents, are less likely to be representative of the population of firms or households.8 While these generic issues are important for all surveys, the focus here is on difficulties specific to measuring governance using survey data.

Some survey questions on governance can be vague and open to interpretation. An interesting example comes from the innovative recent work by Razafindrakoto and Roubaud (2006). They use specially designed surveys in eight African countries to contrast corruption perceptions based on household surveys with those based on expert assessments. The unique feature of this exercise is that the experts were asked to predict the country-level average responses from the household survey. Experts’ ratings were essentially uncorrelated with the household survey responses. The authors conclude that the household surveys capture the “objective reality” of petty corruption and that the experts are just plain wrong.

Their interpretation that there is measurement error only in the expert assessment and not in the household survey is contestable. Households were asked whether they had been “victims of corruption.” There are a variety of reasons why households might falsely think they were victimized by corruption. For example, a patient waiting in line to see a state-provided doctor might think (incorrectly) that people at the head of the line had bribed someone to get there. Conversely, households might well have paid a bribe, received the associated benefit, and found themselves quite satisfied and not at all “victimized” by the transaction. A more modest
interpretation of their finding is that there likely is measurement error in both the household survey and the matching expert assessments. Moreover, in many other cases, expert assessments and household survey responses are strongly correlated across much larger samples of countries.

Well-designed survey questions on corruption have become increasingly specific. For example, questions in the Executive Opinion Survey of the World Economic Forum in some years have asked firms to specifically report the fraction of contract value solicited in bribes on public procurement contracts. Greater attention is also being paid to techniques that enable respondents to report more truthfully to sensitive questions. For example, questions about corruption put to firms are often prefaced by "in your experience, do firms like your own typically pay bribes for . . . ?" Innovative techniques such as randomized response methods are used to protect the confidentiality of individual responses by allowing respondents to "camouflage" their response to sensitive questions by generating some of their responses at random based on the outcome of a coin toss, although these methods have not yet been widely used in large cross-country surveys. A related concern has to do with surveys in authoritarian countries, where respondents might legitimately be fearful of responding truthfully to any question that might be interpreted as critical of the government.

Another potential difficulty in cross-country surveys is cultural bias. It is often argued that because respondents in different countries may have different norms regarding what does or does not constitute corruption, their responses are not comparable across countries. Presumably, however, these cultural biases should not be present in cross-country expert assessments that are deliberately designed to be comparable across countries. Moreover, in many cases it turns out that surveys and expert assessments tend to produce very similar cross-country rankings. Kaufmann, Kraay, and Mastruzzi (2006) document strong correlations between expert assessments and the World Economic Forum's Executive Opinion Survey for six different dimensions of governance. A glance at table 3 provides similar examples: the cross-country correlation between the corruption assessments of World Markets Online, a commercial rating agency, and the Executive Opinion Survey is 0.88. While culture undoubtedly matters in interpreting survey responses across countries, the problem does not appear to be a first-order difficulty.

In short, each type of data has its own strengths and weaknesses. As neither type of respondent is clearly superior for all purposes, it important to rely on a diversity of data sources.

**Should Aggregate or Individual Indicators Be Used?**

Does it make sense to combine individual indicators of governance into aggregate or composite indicators by combining information from multiple sources? Table 1
includes three aggregate indicators, the WGI, the Corruption Perceptions Index (CPI) of Transparency International, and the very recently released Ibrahim Index of African Governance.

The WGI consist of six aggregate indicators of governance covering more than 200 countries and combining cross-country data on governance provided by 30 organizations. The CPI measures only corruption, using a smaller set of data drawn from nine organizations. The WGI control of corruption indicator uses these nine data sources, as well as 13 others not used in the CPI. The Ibrahim Index is an extremely broad collection of a variety of types of governance indicators and several very broad development outcomes, including per capita income, growth, inequality, and poverty. This makes the Ibrahim Index by far the broadest indicator surveyed here, but it also makes it difficult to think of it as a pure governance indicator, because it contains many development outcomes as well. However, three of the five components of the Ibrahim Index—based primarily on subjective governance measures, such as those used by Transparency International and the WGI—correspond more closely to established notions of governance.

*Ubiquitous Measurement Error*

All governance indicators have limitations, which make them imperfect proxies for the concepts they are intended to measure. The presence of measurement error in all governance indicators that this implies is central to the rationale for constructing aggregate indicators. It is useful to distinguish between two broad types of measurement error that affect all types of governance indicators.

Any specific governance indicator will have measurement error relative to the concept it seeks to measure, because of intrinsic measurement challenges. A survey question about corruption, for example, will have the usual sampling error associated with it. Efforts to objectively document the specifics of the institutional environment or regulatory regime face challenges in coming up with a factually accurate description of the relevant laws and regulations in each setting. Measures of the composition and volatility of public spending, for example, which are sometimes interpreted as indicators of undesirable policy instability, are subject to all of the usual difficulties in measuring public spending consistently across countries and over time. Finally, different groups of experts may come up with different assessments of the same phenomenon in a particular country. These divergences of opinion can also be interpreted as measurement error.

To the extent that one is interested in broad concepts of governance, any specific indicator is almost by definition an imperfect measure of the broader concepts to which it pertains, no matter how accurate or reliable it is. A specific assessment of corruption in public procurement would not be fully informative.
about overall corruption in the public sphere even if it were fully accurate about
this specific type of corruption. Information about the statutory requirements for
business entry regulation need not reflect the actual practice of how these
requirements are implemented on the ground, and they are not informative about
regulatory burdens in other areas. Information about freedom of the press is only
one of many factors contributing to the accountability of governments to their
citizens. Notwithstanding some clear advantages that specificity of an indicator
may have for some purposes, one should be careful not to interpret them as suffi-
cient statistics for broader notions of governance.

How important is this measurement error? Unfortunately, the vast majority of
governance indicators do not explicitly acknowledge the extent of measurement
error. One of the few exceptions is the WGI, discussed below. Fortunately, some
simple calculations can shed light on the likely magnitude of measurement error
in individual governance indicators as well. The key to doing so is to identify pairs
of indicators that measure similar concepts, up to an unavoidable measurement
error component. A useful way to interpret the imperfect correlation between the
World Bank’s CPIA and African Development Bank’s CPIA regarding transpar-
ency and corruption is to note that both are measuring the same concept of
transparency and corruption but with a degree of measurement error. Intuitively,
the less measurement error there is in these two sources, the more correlated
they should be. Thus one can interpret the correlation between them as saying
something about the degree of measurement error present.

More formally, think of the observed scores from two organizations, $y_1$ and $y_2$,
as a combination of a signal of unobserved governance, $g$, and source-specific
noise, $e_1$ and $e_2$ (that is, $y_1 = g + e_1$ and $y_2 = g + e_2$). Assume that the variance
of measurement error in the assessments of the two organizations is the same,
and without loss of generality, that the variance of governance is one.\(^{10}\) Some
simple arithmetic reveals that the standard deviation of measurement error is
$SD(e) = \sqrt{(1 - \rho)}/\rho$, where $\rho$ is the correlation between the two expert
assessments.\(^ {11}\) For several pairs of indicators discussed in the article, this standard devi-
ation of error ranges from 0.70 to 1.53 (table 4). The standard errors associated
with standard deviation of error indicators such as the WGI are much smaller,
reflecting the benefits of aggregation in reducing noise in the individual indi-
cators. The standard error for the WGI estimate of control of corruption for a
typical country in 2006 is just 0.17, or less than a quarter of the standard error
of the most precise pair of individual indicators in this example.

To appreciate the magnitude of this measurement error, it is useful to go one
step further and calculate the width of a 90 percent confidence interval for govern-
ance based on any one of these individual indicators and on the additional
assumption that governance and the error term are jointly normally distributed.
The width of the confidence interval is $2 \times 1.64 \times SD(g|y) = 3.28 \times \sqrt{1 - \rho}$.
Table 4. Measurement Error in Individual Governance Indicators

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correlation</th>
<th>Standard deviation of error</th>
<th>Width of confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transparency, accountability, and corruption</td>
<td>0.67</td>
<td>0.70</td>
<td>1.88</td>
</tr>
<tr>
<td>World Bank CPIA-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African Development Bank CPIA-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business entry regulation</td>
<td>0.48</td>
<td>1.04</td>
<td>2.37</td>
</tr>
<tr>
<td>World Economic Forum Executive Opinion Survey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doing Business</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elections</td>
<td>0.30</td>
<td>1.53</td>
<td>2.74</td>
</tr>
<tr>
<td>Global Corruption Barometer Survey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Integrity Elections Index</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors' analysis based on data described in the text.

Since the assumptions imply that 95 percent of countries would have governance levels between -2 and 2, these figures imply that a 90 percent confidence interval for governance for any individual country would span one-half to two-thirds of the entire most likely range of the governance indicator.

Why Aggregate Indicators?

All indicators of governance include measurement error. Aggregate indicators of governance can be a useful way of combining, organizing, and summarizing information from different sources, thereby reducing the influence of measurement error in any individual indicator. Aggregation also allows for the construction of explicit margins of error for both the aggregate indicator and its component individual indicators.

The WGI illustrate how these margins of error can be calculated (box 1). The statistical methodology underpinning the WGI (the unobserved-components model) explicitly assumes that the true level of governance is unobservable and that the observed empirical indicators of governance provide imperfect signals of the fundamentally unobservable concept of governance. This formalizes the notion that all available indicators are imperfect proxies for governance. The estimates of governance that come out of this model are simply the conditional expectation of governance in each country, conditioning on the observed data for each country. Moreover, the unobserved-components model allows one to summarize uncertainty about these estimates for each country with the standard deviation of unobserved governance, conditional on the observed data. These standard deviations can be used to construct confidence intervals for governance estimates, often referred to informally as margins of error. Intuitively, the larger the number of data sources available for a given country, the smaller these
margins of error should be. The variance of the error term can be estimated in each individual underlying governance indicator using this methodology, following a calculation that generalizes the simple one discussed above.

---

**Box 1. The Worldwide Governance Indicators: Critiques and Responses**

The Worldwide Governance Indicators (WGI) are among the most widely used cross-country governance indicators (see Kaufmann, Kraay, and Mastruzzi 2007a for a description). They report on six dimensions of governance for more than 200 countries for 1996–2006. The indicators are based on hundreds of underlying individual indicators drawn from 30 organizations, based on responses from tens of thousands of citizens, enterprise managers, and experts.

As one of the most prominent and widely used collections of cross-country governance indicators, the WGI have naturally generated criticism. Most of these criticisms appear largely invalid (Kaufmann, Kraay, and Mastruzzi 2007b).

**Lack of Comparability over Time and across Countries**

Several critics have raised concerns that the WGI are not comparable over time and across countries, that the indicators use units that set the global average of governance to be identical in all periods, that comparisons of pairs of countries or single countries over time are based on different sets of underlying data sources, and that there are substantial margins of error in the aggregate WGI.

These criticisms appear unjustified for several reasons. First, there is no clear evidence of a trend in one direction or another in global averages of governance in any of the underlying individual data sources (the overall evidence pointing to general stagnation). The choice of a constant global average is therefore no more than an innocuous choice of units. Second, changes in the set of underlying data sources on average contribute only minimally to changes over time in countries’ scores on the aggregate WGI: the majority of cross-country comparisons using the aggregate WGI are based on a substantial number of common data sources. Third, the presence of explicit margins of error in the WGI is an important advantage, serving as a useful antidote to superficial comparisons of country ranks or country performance over time that are often made with other governance indicators. A substantial fraction of cross-country and over-time comparisons using the WGI result in statistically significant differences, suggesting that the WGI are informative.

**Biases in Expert Assessments**

Several critics have alleged various biases in the data sources underlying the WGI, including an excessive emphasis on business-friendly regulation on the part of some data providers; ideological biases, such as a bias against left-wing governments, on the part of some data providers; and “halo effects,” whereby countries with good economic performance receive better-than-warranted governance scores. Convincing empirical evidence in support of such biases has not been provided. Empirical work by Kaufmann, Kraay, and Mastruzzi (discussed in the main text) suggests that these biases are quantitatively unimportant.

**Correlated Perception Errors**

Several critics have suggested that expert assessments make similar errors when assessing the same country, leading to correlations in the perception errors across expert assessments. While this is plausible, there is little convincing empirical evidence: work by Kaufmann, Kraay, and Mastruzzi (discussed below) suggests that these biases are quantitatively unimportant.

(Continued)
Box 1. Continued

A related concern is that correlated perception errors will lead to the overweighting of such sources in the aggregate WGI, which weights individual data sources by estimates of their precision, which in turn are based on the observed intercorrelation among sources. Given the at best modest evidence of correlated perceptions errors, this is unlikely to be quantitatively important. The WGI country rankings are highly robust to alternative weighting schemes (Kaufmann, Kraay, and Mastruzzi 2006).

Definitional Issues

Some critics have taken issue with the definitions of governance and thus the assignment of individual governance indicators to the six aggregate WGI. As there is no consensus on the definition of governance, there cannot be any right or wrong definitions or corresponding measures of governance. That said, most reasonable definitions of governance cover similar broad areas, and aggregate indicators capturing these broad areas are likely to be similar. Moreover, as virtually all of the individual indicators underlying the WGI are publicly available on the WGI Web site, researchers can easily construct alternative indicators corresponding to their preferred notions of governance.

Reliance on “Subjective” Data

Various critics have argued that the perceptions-based data on which the WGI are based do no more than reflect vague and generic perceptions rather than specific objective realities and that “specific, objective, and actionable” measures of governance are needed to make progress in governance reforms. Virtually all governance indicators necessarily involve some element of subjectivity. Perceptions-based data are extremely valuable, because they capture the views of relevant stakeholders who act on these views. Moreover, the links from specific changes to policy rules are very difficult to link to changes in outcomes of interest, making it difficult to identify indicators that are “action worthy” as opposed to merely “actionable.”

From the standpoint of users the margins of error associated with estimates of governance are nontrivial. For many pairs of countries with similar scores on the 2006 WGI control of corruption indicator, the confidence intervals overlap, indicating that the small differences between them are unlikely to be statistically, or practically, significant (figure 2). However, possible pair-wise comparisons between countries do result in significant differences. Roughly two-thirds of the possible pair-wise comparisons of corruption across countries result in differences that are significant at the 90 percent confidence level, and nearly three-quarters of comparisons are significant at the 75 percent confidence level. Clearly, far fewer pair-wise comparisons would be significant if they were based on any single individual indicator, whose margins of error had not been reduced by averaging across alternative data sources. For the WGI control of corruption indicator, for example, only 16 percent of cross-country comparisons based on a typical individual data source, such as Global Insight-DRI, would be significant at the 90 percent confidence level.

The WGI are unusual among governance indicators in their transparent recognition of such margins of error. The vast majority of investment climate and governance indicators simply report country scores or ranks, without quantifying the
measurement error these rankings inevitably contain. This has contributed to a spurious sense of precision among users of these indicators and to an overemphasis on small differences across countries.

Of course, aggregate indicators have their own shortcomings. Foremost among them is the inevitable loss of specificity. Averaging one indicator of judicial corruption and another indicator of bureaucratic corruption arguably yields a more informative indicator of overall corruption, but not necessarily a more informative indicator of either of the two specific types of corruption. Averaging an indicator of freedom of the press with an indicator of electoral integrity yields a more informative indicator of overall democratic accountability, but not of either particular concept. For some purposes the broad aggregate indicators will be useful; for others the disaggregated indicators will be more useful. This is not a shortcoming, however, because virtually all aggregate governance indicators can readily be disaggregated into their constituent components, giving the user the freedom to choose the appropriate level of aggregation for the task at hand.
The second concern with aggregate indicators is that their effectiveness at reducing measurement error depends crucially on the extent to which their underlying sources provide independent information on governance. Some types of expert assessments may make correlated errors in their governance rankings (although empirical evidence suggests that these error correlations are not likely to be very large). Aggregate indicators can mitigate only the component of measurement error that is truly independent across the different underlying indicators. This point is particularly relevant when contrasting multiple- and single-source aggregate indicators.

The WGI are multiple-source aggregate indicators, combining information from a large number of sources. In contrast, many other data sources report aggregates of their own subcomponents. For example, there is an aggregate CPIA rating in conjunction with the 16 underlying components, and there are six aggregate Global Integrity indicators, which combine information from more than 200 underlying individual indicators. All of the underlying individual indicators for a given country are scored by the same respondents. As a result, any respondent-specific biases are likely to be reflected in all of the individual indicators; the gain in precision from relying on the aggregate indicators from these sources will not be as large as when aggregate indicators are based on multiple sources.

In summary, aggregate governance indicators can play a useful role in synthesizing and summarizing the large variety of individual governance indicators. Using aggregate indicators is one way to exploit the complementarities between the different types of indicators (rules or outcomes, surveys or experts). Aggregation can also increase the precision with which these aggregate indicators measure broad but unobservable concepts of governance. Of course, for some purposes, more specific indicators are useful. It is thus important to be able to easily disaggregate aggregate indicators into their constituent components, as is the case with the WGI.

**Moving Forward**

A sobering picture emerges from this review: while most indicators of governance have many virtues, all face distinct challenges. Researchers, therefore, need to look at a variety of indicators and sources when monitoring or assessing governance across countries, within a country, or over time. A few principles may be useful as this work, as the use of governance indicators in public sector policy-making and civil society monitoring, continues.

Avoid false dichotomies. Too often, discussions of governance indicators over-emphasize distinctions between types of governance indicators, with insufficient regard for the strong complementarities between them. Artificially, sharp
distinctions are often drawn between “subjective” and “objective” indicators of governance, when, in fact, virtually all indicators of governance rely on the judgments or perceptions of respondents in one way or another. In some cases, even the terminology is misleading. The recently released Ibrahim Index of African Governance, for example, touts itself as providing objective assessments of governance, even though its core governance components are based primarily on purely subjective data, including the Transparency International CPI and subjective ratings by the Heritage Foundation and the Economist Intelligence Unit.

Distinctions between aggregated and disaggregated indicators often have an artificial element also. Some aggregate indicators transparently disclose each disaggregated source, enabling users to take advantage of the complementarities between the two types of indicators and blurring the distinction between the two. For some purposes, it is useful to combine information from many individual indicators into some kind of summary statistics, while, for other purposes, the disaggregated data are of primary interest. Even where disaggregated data are of primary interest, however, it is important to rely on a number of independent sources for validation, because the margins of error and the likelihood of extreme outliers are significantly higher for a disaggregated indicator.

An excessively narrow emphasis on “actionable” indicators detailing specific policy interventions immediately under the control of governments can divert attention from equally important discussions of which of these indicators are “action worthy,” in the sense of having significant impacts on outcomes of interest. The answer is often context-specific and rarely obvious a priori. Focusing too much on “actionable” indicators while downplaying scrutiny of outcome indicators may result in undue emphasis on measures that may not translate into concrete progress.

Use indicators appropriate for the task at hand. As with all tools, different types of indicators are suited for different purposes. Governance indicators can be used for regular cross-country comparisons. While many of these indicators have become increasingly specific, they often remain blunt tools for monitoring governance and studying the causes and consequences of good governance at the country level. For these purposes, a wide variety of innovative tools and methods of analysis has been deployed in many countries (reviewing these methods is beyond the scope of this survey). Examples of in-country tools include the World Bank’s Investment Climate Assessments, the World Bank Institute’s Governance and Anti-Corruption diagnostics, the corruption surveys conducted by some chapters of Transparency International, and the institutional scorecard carried out by the Public Affairs Center in Bangalore, India. Many project-specific interventions and diagnostics are possible to measure governance at this level.¹³

Public and professional scrutiny is essential for the credibility of governance indicators. Virtually all of the governance indicators listed in table 2 are publicly
available, either commercially or at no cost to users. This transparent feature is central to their credibility for monitoring governance. Open availability permits broad scrutiny and public debate about the content and methodology of indicators and their implications for individual countries. Many indicators are also produced by nongovernmental actors, making it more likely that they are immune from either the perception or the reality of self-interested manipulation on the part of the government. Scholarly peer review can also strengthen the quality and credibility of governance indicators. For example, articles describing the methodology of the Doing Business indicators, the Database of Political Institutions, and the WGI have appeared in peer-reviewed professional journals. Transparency with respect to details of methodology and its limitation is also essential for credible use of governance indicators. It is important that users of governance indicators understand fully the characteristics of the indicators they are using, including any methodological changes over time and time lags between the collection of data and publication.

It is thus of concern that some proposed and existing indicators of governance are insufficiently open to public scrutiny. While the recent disclosure of the World Bank’s CPIA ratings for low-income countries represents a positive step, these indicators are being disclosed for only about half of the roughly 130 countries for which they are prepared each year, and none of the historical data from 2005 or earlier are publicly available. Historical data on the CPIA ratings of the African Development Bank and Asian Development Bank have also not been disclosed publicly. This is unfortunate, given that the decision to selectively disclose recent CPIA data and not to disclose historical CPIA data is made by the executive boards of these organizations and therefore reflects the desire of the very governments these ratings are supposed to assess. Regarding transparency, it is also of concern that although the Public Expenditure and Financial Accountability initiative has been ongoing since 2000, it had resulted in indicators and reports on just 42 countries as of March 2007, for only one period per country, only nine of them publicly available. Moreover, because these reports are prepared in collaboration with the governments in question, their credibility may not be the same as those associated with third-party indicators. Similar concerns affect recent Organization for Economic Co-operation and Development-led efforts to construct indicators of public procurement practices.

Transparency acknowledge margins of error of all governance indicators. All governance indicators include measurement error and so should be thought of as imperfect proxies for the fundamentals of good governance. This is not just an abstract statistical point, but rather one of fundamental importance for all users of governance indicators. Wherever possible, such margins of error should be explicitly acknowledged, as they are in the WGI, and taken seriously when the indicators are used to monitor progress on governance. At times the lack of disclosure of
margins of error is rationalized by suggesting that they would be missed by most readers. Experience with the WGI suggests that this is not the case, with many users recognizing and benefiting from this additional degree of transparency about data limitations.

*Exploit the wealth of available indicators, recognizing that progress in developing new indicators is likely to be incremental.* Much more work needs to be done to exploit the large body of disaggregated measures of governance already in existence. Linking disaggregated indicators to disaggregated outcomes, both across countries and over time, is likely to be an important area of research over the next several years that is likely to have important implications for policymakers.

There is also scope for developing new and better indicators of governance. Work to improve such indicators will be important, as indicators are increasingly used to monitor the success and failure of governance reform efforts. But given the many challenges of measuring governance, it is important to recognize that progress in this area over the next several years is likely to be incremental rather than fundamental. Alongside efforts to develop new indicators, there is also a case to improve existing indicators, particularly in increasing the periodicity of heretofore one-off efforts and in broadening their country coverage (covering industrial and developing economies), as well as covering issues for which data are still scarce, such as money laundering.

Notes

Daniel Kaufmann is a director of global programs at the World Bank Institute; his email address is dkaufmann@worldbank.org. Aart Kraay is a lead economist in the Development Research Group at the World Bank; his email address is akraay@worldbank.org. The authors would like to thank Shanta Devarajan for encouraging them to write this survey. Simeon Djankov and three anonymous referees for their helpful comments, and Massimo Mastruzzi for assistance.

1. For surveys of and user guides to governance indicators, see UNDP (2005), Arndt and Oman (2006), and Knack (2006). Because of space constraints, no attempt is made here to review the important body of work focused on in-depth within-country diagnostic measures of governance that are not designed for cross-country replicability and comparisons.


3. Indeed, this is reflected in the terminology of “actionable” governance indicators emphasized in the World Bank’s Global Monitoring Report (World Bank 2006).

4. See King and Wand (2007) for a description of how this problem can be mitigated by the use of “anchoring vignettes” that provide a common frame of reference to respondents in interpreting the response scale. The basic idea is to provide an understandable anecdote or vignette describing the situation faced by a hypothetical respondent to the survey. For example, “Miguel frequently finds that his applications to renew a business license are rejected or delayed unless they are accompanied by an additional payment of 1,000 pesos beyond the stated license fee.” Respondents are then asked to assess how great corruption as an obstacle is for Miguel’s business, using a 10-point scale. Since all respondents use the scale to assess the same situation, this rating can be used to “anchor” their
responses to questions referring to their own situation.

5. These two indicators are measured as the average of 14 “in law” components and the 20 “in practice” components of the elections indicator of Global Integrity.

6. Starting with the 2005 data, both the African Development Bank and the World Bank have made their CPIA scores public. The African Development Bank does so for all borrowing countries; the World Bank does so only for countries eligible for its most concessional lending.

7. Kaufmann, Kraay, and Zoido-Lobatón (1999a) show how the estimated margins of error of their aggregate governance indicators would increase if they assume that the error terms made by individual data sources were correlated. Recently, Svensson (2005), Arndt and Oman (2006), and Knack (2006) have raised this criticism again, largely without the benefit of systematic evidence. Kaufmann, Kraay, and Mastruzzi (2007b) provide a detailed response.

8. This is not to say that all of the surveys used to measure governance are necessarily representative in any strict sense of the term. In fact, one general critique is that several large cross-country surveys of firms that provide data on governance are not very clear about their sample frame and sampling methodology. The Executive Opinion Survey of the World Economic Forum, for example, states that it seeks to ensure that the sample of respondents is representative of the sectoral and size distribution of firms (World Economic Forum 2006). But it reports that it “carefully select[s] companies whose size and scope of activities guarantee that their executives benefit from international exposure” (p. 133). It is not clear from their documentation how these two conflicting objectives are reconciled.

9. A simple example is that respondents are asked whether they have ever offered a bribe. But before answering, the respondent is instructed to privately toss a coin and to answer “yes” if either they have in fact offered a bribe, or the coin comes up heads. See Ahsan and Murrell (2006) for an assessment of the extent to which randomized response methods correct for respondent reticence and an innovative approach to using this methodology to weed out less than candid respondents.

10. The assumption of a common error variance is necessary in this simple example with two indicators in order to achieve identification. In this example, just one sample correlation in the data can be used to infer the variance of measurement error; just one measurement error variance can thus be identified. In more general applications of the unobserved components model, such as the WGI, this restriction is not required because there are three or more data sources.

11. For details on this calculation, see Kaufmann, Kraay, and Mastruzzi (2004, 2006). Gelb, Ngo, and Ye (2004) perform a similar calculation comparing the African Development Bank and World Bank CPIA scores. Their conclusion that the CPIA ratings have little measurement error is driven largely by the fact that the authors focus on the aggregate CPIA scores, which are very highly correlated between the two institutions. The focus here is on one of 16 specific questions; at this level of disaggregation, the correlation between the two sets of ratings is considerably lower.

12. For example, virtually all of the individual indicators underlying the aggregate WGI are available at www.govindicators.org.

13. One of the best-known and best-executed recent studies of this type is a study of corruption in a local road-building project by Olken (2007).

References


Two Comments on "Governance Indicators: Where Are We, Where Should We Be Going?" by Daniel Kaufmann and Aart Kraay

The following comments by Shantayanan Devarajan and Simon Johnson provide two perspectives on indicators in general and the World Governance Indicators in particular.

Shantayanan Devarajan

The World Bank Research Observer publishes balanced surveys of the literature. When the authors of a survey are also the proponents of one of the major indicators being surveyed, it invites comments to ensure that balance is maintained.

Kaufmann and Kraay provide a useful taxonomy of governance indicators, distinguishing between those measuring "rules on the books" and "rules on the ground" and those reflecting the views of experts and the results of surveys. While providing a balanced overview of the pros and cons of different methods, they make a strong case in favor of measuring rules on the ground based on an aggregated mix of expert- and survey-based indicators, along the lines of their World Governance Indicators (WGI).

Any assessment of governance indicators—or any indicator for that matter—must be based on the purposes to which the indicators will be put, as Kaufmann and Kraay note. This comment examines how well the WGI and other indicators perform in two specific instances.

The first is the allocation of resources—such as the concessional aid provided by the World Bank’s International Development Association (IDA) or the US Millennium Challenge Corporation (MCC)—across low-income economies. These resources are allocated according to a formula that includes, among other factors, the productivity of aid in reducing poverty in a particular country—a factor that is partly a function of the quality of governance. An indicator is thus needed that...
measures the quality of governance across countries. By focusing on rules on the
ground, using a mix of expert- and survey-based information, aggregating across
indicators within a country, and providing standard errors around the means
of these aggregate indicators, the WGI provide a defensible method of making
cross-country comparisons. The IDA uses the judgment of World Bank experts in
assessing a country’s policies and institutions, but that judgment is informed by
the WGI. The MCC uses five of the WGI, along with 11 other indicators, to deter-
mine a country’s eligibility for its programs.

The WGI have produced some seemingly anomalous results (India, for instance,
ranks in the bottom quartile worldwide on “political stability and lack of vio-
lence”). But aggregating across surveys in each country is needed if results are
not to reflect surveys conducted across only a limited set of countries or some
other bias-inducing method.

A second possible use of the WGI is to help identify the nature of the “govern-
ance problem” in a country. Bangladesh has ranked in the bottom quartile world-
wide (and considerably below the low-income country average) in all but one or
two of the aggregate WGI for the past decade and has been at or very near the
bottom of Transparency International’s corruption indicators. At the same time,
the growth rate of per capita income has risen 0.1 percent every decade since the
1970s (the growth rate is now close to 5 percent), and poverty fell 0.8 percent in
five years (twice the rate in India). The country has already achieved universal
primary enrollment and gender equity in secondary school, and it is on track to
reduce child mortality by two-thirds (relative to 1990 levels) by 2015.

In what sense, then, does Bangladesh have a “governance problem?” Is it poss-
ible that the data on development indicators have been mismeasured? Although
some development indicators may be weaker than they appear—there is some
evidence, for example, that enrollment rates among the lowest quintile of
Bangladesh’s population remain extremely low (World Bank 2007)—the country
has clearly made tremendous gains in development over the past two decades. It
is possible that it would have performed even better absent its governance problems
(World Bank 2006). But, this does not answer the question of how Bangladesh’s
development outcomes are so much better than those of other countries with
“better” governance.

One explanation is that governance indicators in Bangladesh fail to capture
the fact that the country has a vibrant and active civil society that not only deli-
vers services, but provides some accountability to government. The WGI also
seem to overlook the increasingly mature media, including vernacular newspa-
pers, which play something of a watchdog role. But these measurement errors
may be second-order considerations compared with the fundamental facts that
the governance indicators are capturing. Bangladesh does score in the 30th per-
centile on the WGI voice and accountability indicator; it comes close to the
25th percentile on government effectiveness. But it ranks in the 5th percentile on control of corruption. These rankings are very low for a country that is performing so well.

Another explanation is that the relation between governance and development in Bangladesh is unique, because the Bangladeshi people have worked around the country's governance problems to spur development. When the country was born, out of a civil war, there was hardly a government. International and national nongovernmental organizations (NGOs) filled the vacuum by delivering basic services, such as health, family planning, and education, and by creating microcredit schemes. As these efforts proved effective, the government made space for these NGOs and the private sector, in some cases contributing to their financing. The government funds secondary education, for example, although 95 percent of it is provided by the nonstate sector. Similarly, Bangladesh's garment export sector grew rapidly, thanks to duty drawback systems and bonded warehouses that enabled textiles to come into the country duty-free, circumventing a highly opaque customs system.

Of course, this explanation does not explain why Bangladesh was able to work around its governance problems when so many other countries were not. Bangladesh is a densely populated, homogeneous society, in which innovations spread like wildfire. Soon after one village discovers something that works, neighboring villages find out about it and adopt it. As a result, family planning, microfinance, and other programs took off in Bangladesh more easily than they might have elsewhere. By the time the government grew strong enough to control the NGOs and others, it was too late, microfinance, family planning, and private schooling had already become commonplace. (To its credit, the government recognized this and proceeded to support the providers of essential services with finance.) The result is a country with weak governance indicators, but impressive development.

The Bangladesh case illustrates the fact that governance indicators such as the WGI do not capture the multifaceted ways in which governance affects development in a particular country. It would be dangerous to use indicators to jump to simple conclusions without understanding the specific relation between governance and development in a particular country; the indicators should certainly not be used by themselves to design policy responses to problems of weak governance.

This is the downside of having an indicator that permits intercountry comparisons of governance: the richness of country-specific detail is lost. Kaufmann and Kraay recognize this tradeoff in their concluding discussion about more disaggregated indicators. One should not go too far down this path, however, lest the main benefit of the WGI and other such indicators—their comparability across countries—be diluted in the quest for indicators that are more country specific. In
the Tinbergen tradition of not having more objectives than instruments, one should not expect governance indicators to serve too many purposes.

Notes

Shantayanan Devarajan is the chief economist of the South Asia Region at the World Bank; his email address is sdevarajan@worldbank.org.

1. This result may not be as anomalous as it appears at first blush; however, according to the Indian prime minister, 170 of 602 districts have a significant Naxalite (Maoist) presence (Singh 2005).

References

Singh, Manmohan 2005. “PM’s Reply in Rajya Sabha to the Debate on Motion of Thanks to the President’s Address.” March 11. New Delhi.


Simon Johnson

Kaufmann and Kraay took a major step forward in thinking about governance when (together with Pablo Zoido-Lobatón) they first published their World Governance Indicators (WGI), in 1999. These indicators provided a new way to combine comparable indices along various dimensions of governance. The WGI included more countries and more pieces of data than had previously been available, complementing various other measures.

From the beginning, the WGI have clearly indicated the underlying sources of the data and how the data are constructed, explicitly reporting error bands for all estimates. This was a major innovation and remains of paramount importance.

The authors (together with Massimo Mastruzzi) have also done a great service by continuing to update the data and refine their methods. The WGI are now well established as one of the standard sets of measures that any researcher or policy analyst must consult. No measure is perfect, of course, but anyone trying to establish that a particular set of results is or is not robust is well advised to work carefully through the various vintages of the WGI.
The appearance of the WGI coincided with a major backlash against carefully measuring and thinking hard about governance. While this backlash initially seemed rather academic and sometimes arcane, it had considerable immediate impact on both policy discussions and the impetus to research in this area. Rather than building on the WGI, serious endeavor in this area lost momentum, as three main arguments against governance indicators gained currency.

The first is that because the WGI explicitly report “errors,” they cannot be relied on. This argument is completely mistaken: similar error bands should be reported for all the data used in economics, particularly for countries with weak statistical systems.

The second argument is that governance measures are not useful because many countries are growing, despite weak governance. What are the uses of these measures if they cannot predict who will and will not grow? While countries with weak institutions can, and indeed often do, grow for prolonged periods, over the longer run, it is hard to escape weak governance. Far from the exception, growth spurts are actually a standard feature almost everywhere (Berg, Ostry, and Zettelmeyer 2006)—as is the inability of many countries with weak governance to sustain the gains made during the good years.

Some countries—albeit only relatively few recently—have managed to sustain rapid growth, despite weak initial governance. Too little is known about these cases, but one common element appears to be a strong focus on exports, particularly of manufacturing goods (see Johnson, Ostry, and Subramanian 2007 for specific country examples).

These exceptions notwithstanding, if a country’s governance indicators are weak—as measured by the WGI—the presumption should be that sustained growth will be difficult. Strengthening governance should help increase the odds of sustaining growth.

The third and the most interesting argument is that aggregate indicators miss a great deal of rich detail, failing to pick up the de facto arrangements that effectively take the place of more formal governance rules. Close observation of how society is organized always turns up functional forms of improvisation—that is, ways to organize transactions even where contract enforcement is weak. Some informal mechanisms may be quite efficient, in the sense that the transaction costs are not much higher than they would have been had formal contracting worked well.

Some of this organization can occur in and around the manufactured export sector. Initial improvisation could lead to more durable rules over time, both in the manufacturing sector and, through institutional spillovers, in the country as a whole.

One needs to be very careful about drawing policy implications from apparent conditions that cannot be measured; however, it is hard to make much progress.
on the basis of anecdotal evidence. Informal governance arrangements may shift and be fragile; they may also prevent people from entering new activities. The limited work on this phenomenon suggests that informal governance is often a weak substitute for well-functioning governance.

What is needed is a WGI-type approach to measure governance at a more local level, for both sectors and cities. There have already been ambitious steps in this direction [for example, Kaufmann, Leautier, and Mastruzzi (2005)], and there has been some indication that the World Bank, which is uniquely positioned in this regard, would move in this direction. So far, however, progress has been limited.

There are many good reasons to move carefully in this area, and no doubt we need to think carefully about how to make best use of resources—something that Kaufmann and Kraay reflect in this article, with their balanced assessment of alternative sources within and outside the WGI. But it would be unfortunate if the legitimate debate over interpretation of macrolevel governance measures were to undermine confidence in this area of enquiry and prevent further work from being undertaken. The field is still at an early stage. Better measures are needed to gauge what works and what does not. Such measures should stand firmly on the shoulders of the WGI.

Note

Simon Johnson is an economic counselor and director of the Research Department at the International Monetary Fund; his email address is sjohnson@imf.org.

References


When growth-promoting spending is cut so much that the present value of future government revenues falls by more than the immediate improvement in the cash deficit, fiscal adjustment becomes like walking up the down escalator. Although short-term cash flows matter, too tight a focus on them encourages governments to invest too little. Cash-flow targets also encourage governments to shift investment spending off budget by seeking private investment in public projects, irrespective of its real fiscal or economic benefits. To deal with this problem, some observers have suggested excluding certain investments (such as those undertaken by public enterprises deemed commercial or financed by multilaterals) from cash-flow targets. These stopgap remedies may help protect some investments, but they do not provide a satisfactory solution to the underlying problem. Governments can more effectively reduce the biases created by the focus on short-term cash flows by developing indicators of the long-term fiscal effects of their decisions, including accounting and economic measures of net worth, and, where appropriate, including such measures in fiscal targets or even fiscal rules. JEL codes: O23 E62 H60 H54

A popular phrase during the era of macroeconomic stabilization of the 1990s was “adjustment with growth.” The focus of this article is on the surprising possibility that some types of fiscal austerity not only fail to bring growth, but they may not even bring “adjustment” in the long run.

Consider the following anecdote from the World Bank’s own budgeting experience. In 1993 the World Bank Research Department unexpectedly produced a bestseller entitled The East Asia Miracle. The Research Department soon exhausted its administrative budget allocation for reprinting the book. The World Bank’s centralized budget department denied a request for extra budgetary resources for printing more copies of the book on the grounds that the Research Department...
had already exceeded its printing budget—even though producing more copies of
the book would have more than paid for itself!

This kind of unreason is not confined to the world of bureaucratic budgetary
management; it also extends to fiscal policy practice. The primary concern of
most fiscal programs is to ensure public sector solvency, commonly viewed as an
essential ingredient of macroeconomic stability. Solvency is by definition an inter-
temporal concept, relating to the present value of revenues and expenditures and
encompassing both assets and liabilities. A cut in public investment that lowers
growth will lower the present value of revenues; it is conceivable that the govern-
ment’s intertemporal position deteriorates at the same time as the cash deficit
improves. In practice, however, it is customary to assess the strength of public
finances almost exclusively on the basis of the cash deficit (or “overall balance”)—
that is, the rate of acquisition of debt by the public sector.

Latin America offers a good illustration of this practice. There is rising concern
across the region that the fiscal adjustment that many countries had to undertake
since the early 1990s may have come with an excessive fall in public investment
(figure 1). To the extent that the response of private investment has been insuffi-
cient to offset the decline of public investment in key sectors, such as infrastruc-
ture, current levels of public investment are perceived by many as too low to
support long-term growth rates consistent with rapid poverty reduction.

Political economy considerations make adjustment difficult at the best of times.
The recent backlash against free-market reforms in Latin America and the long-
standing sensitivity to conditions perceived as imposed by outsiders make it more
important than ever that adjustment programs be well conceived.

**Figure 1.** Primary Deficit and Public Infrastructure Investment in Latin America, 1980–2001

![Primary Deficit and Public Infrastructure Investment in Latin America, 1980–2001](image)

*Note:* Figure is based on data from Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Mexico, and Peru.

The international evidence suggests that Latin America's experience is the rule rather than the exception. Declines in infrastructure spending often account for the lion's share of fiscal deficit reduction, as Hicks (1991) shows for developing economies in a cross-regional context. Easterly and Servén (2003) show for Latin America, and Estache (2004) shows for Sub-Saharan Africa. For industrial countries, Roubini and Sachs (1989) and De Haan et al. (1996) find that capital expenditure falls disproportionately at times of fiscal stringency. Balassone and Franco (2000) show that fulfillment of the Maastricht deficit targets sped the decline of public investment in the European Union (figure 2): of the nine countries that exceeded the deficit target in 1992, eight met it in 1997. In all eight, public investment had fallen relative to GDP; in seven of them, it had also fallen relative to total primary expenditure. In contrast, three of the six countries that met the target in 1992 raised their public investment in the subsequent years.

The tendency toward compression of public investment at times of fiscal austerity underlies the fact that investment is the most volatile of all public spending items, as Talvi and Végh (2000) document using data from developing economies and Lane (2003) documents using data from industrial countries. Of course, declining public investment would be of little consequence if it reflected improved spending efficiency or were fully matched by increased investment by the private sector. In most countries in Latin America, the only developing region for which adequate data are available, this may have been the case in the telecommunications sector. But the evidence suggests that in most infrastructure sectors in most countries, private investment did not offset public sector retrenchment (Calderón and Servén 2004).

Declining investment is a cause for concern when it results in decreased accumulation of public capital and public capital is productive. This is not always

---

**Figure 2.** Primary Deficit and Public Investment in the European Union, 1980–2002

![Figure 2](image)

*Note: Figure is based on data from Austria, Belgium, Finland, France, Germany, Spain, Sweden, and the United Kingdom.*

*Source: OECD Economic Outlook database.*
the case; many projects labeled as public investment can be white elephants, which bring no future output benefits. The link between public investment spending and capital accumulation can be fragile if investment involves significant waste—when projects are poorly selected and public procurement is inefficient or beset by corruption, for example (Pritchett 2000). With weak governance, public investment may become a vehicle for dispensing political favors rather than acquiring productive assets (Keefer and Knack 2007).

The empirical literature is far from unanimous on the contribution of public capital to aggregate output or growth; this lack of agreement is hardly surprising in the context of growth empirics. Nevertheless, most studies, especially the more recent ones, do find a positive impact. The conclusions appear to depend in part on the approach followed: studies using measures of physical infrastructure assets find significantly positive output contributions in the vast majority of instances, whereas those that measure public capital using cumulative investment flows tend to be less conclusive, likely for the reasons outlined in the preceding paragraph. In some cases, however, both approaches yield similar results; using both financial and physical measures of public capita, for example, Ferreira and Araujo (2005) find significantly positive output effects in Brazil.

Moreover, even if wasteful public investment spending weakens the link between spending and outcomes, an across-the-board reduction in public investment will still result in cuts in productive infrastructure projects. Sacrificing such projects weakens the economy’s growth potential; the right response is instead to protect high-return projects from spending cuts. If government does otherwise, it is trying to walk up the down escalator.

This article offers a selective overview of these issues. It draws from the Latin American experience, because it has been relatively well documented, but its concerns are much more general. Indeed, they cut across developing as well as industrial regions, although they are more pressing in developing countries, which still have a long way to go in building up their infrastructure capital stocks. The article draws policy implications for the design and monitoring of fiscal targets consistent with both solvency and the efficient utilization of fiscal resources.

The article is organized as follows. The next section reviews the shortcomings of the current approach to fiscal discipline. The second and third sections deal with two types of remedies: granting exceptions to existing fiscal targets and introducing new targets. The last section offers some concluding comments.

Shortcomings of the Standard Approach to Fiscal Discipline

Fiscal adjustment programs typically focus on the short-term time path of the government’s cash deficit, whose measurement is usually the center of attention
of fiscal accounting. Short-term cash deficits and debt are the key fiscal concern of official creditors and form the basis of loan conditions in the fiscal and macroeconomic dimensions. They are also closely scrutinized by multilateral institutions, private creditors and investors, and economic analysts.

There are good reasons why these fiscal aggregates should be closely watched. The cash deficit approximates the government’s financing needs, which are a primary concern for the fiscal authorities as well as financial market participants. It can also give an indication of the public sector’s contribution to overall aggregate demand and thus its stance from the viewpoint of short-term stabilization, although the primary deficit (which excludes interest payments) may be preferable for this purpose.

Debt and the cash deficit can be misleading as solvency measures, however, because they do not take into account the assets and future income the government may acquire by incurring debt today. This, of course, is hardly surprising: liquidity and solvency are fundamentally different concepts; as in corporate finance, different indicators are needed to gauge them. A corporation does not seek to maximize just this year’s cash flow; it seeks to maximize the present discounted value of all future cash flows. Telling the public sector to improve the cash balance no matter what would be like telling Apple to forgo investing in a new iPod factory in order to improve this year’s cash flow.

Solvency assessments based on debt and the cash deficit implicitly treat all public expenditures in the same way, because they all pose the same claim on today’s fiscal resources. This blurs the distinction between public investment and public consumption and, more precisely, between expenditures that yield future fiscal benefits and those that do not—even though they may have radically different implications for tomorrow’s public revenues and therefore for solvency itself.

Such practice distorts the tradeoffs faced by fiscal policy, both across time and among different kinds of public expenditures. Across time, binding debt and cash deficit targets today tend to encourage postponement of expenditures and advancement of tomorrow’s revenues, even if their present value, which is the relevant concern for solvency, remains unchanged (or declines as a result of delaying urgently needed expenditures, for example). Across expenditure types, liquidity targets pose a one-for-one tradeoff at the margin, regardless of the type of expenditures involved, whereas solvency targets do not. Faced with these tradeoffs, governments having to strengthen public finances frequently choose adjustment paths that, by altering the time profile, the composition of expenditures, or both, attain the prescribed liquidity targets without any significant improvement in solvency. They resort to deferring payments to the first day of the next year, accumulating arrears to government workers or suppliers, advancing the collection of taxes, awarding higher pensions instead of increased wages, or granting guarantees instead of subsidies. Easterly (1999) and Easterly and Servén (2003) provide a variety of examples of this kind of illusory fiscal adjustment.

Easterly, Irwin, and Servén
Thus, other things being equal, governments facing binding liquidity targets today may devote too few resources to expenditures that yield returns tomorrow. This effect of liquidity targets on public spending composition is additional to the biases introduced by other political economy factors. These factors (which include governments' short-time horizons and political clientelism) can distort spending choices by discouraging public expenditures whose benefits accrue in the future in favor of those with immediate fiscal or political payoffs. Far from correcting these distortions, the conventional approach to fiscal discipline magnifies them.

If fiscal adjustment disproportionately cuts infrastructure spending that enhances growth, it can lead to a vicious circle in which low growth generates unsustainable debt dynamics, which force fiscal adjustment implemented through investment cuts, which lowers growth further and prompts additional fiscal retrenchment and investment cuts. In other words, if debt stabilization is pursued primarily by cutting productive spending, destabilization can ensue.

This phenomenon has been documented in both industrial and developing countries. In industrial countries, Alesina and Perotti (1997) find that fiscal corrections based primarily on public investment contraction are typically unsuccessful: they have an adverse effect on growth, and their stabilizing effect on public finances is eventually reversed. Calderón and Servén (2003) review the impact of public investment cuts in selected Latin American countries over the past decade. Their calculations suggest that the ensuing slump in growth and tax collection may have greatly weakened the intended solvency-enhancing effects of the capital expenditure decline.

These issues concern all kinds of public expenditures that generate future fiscal benefits. Public infrastructure investment is the leading example, to the extent that public capital yields financial returns that the government can capture. Conceptually, infrastructure investments can be divided into three groups:

- Investment that generates direct financial returns through user fees, such as ports, airports, railways, and toll roads.
- Investment that does not generate user fees but increases growth and future tax collection.
- Investment that generates no future fiscal or growth benefits, whether or not the project has a positive social return (as in the case of environmental projects).

The first two types of projects may pay for themselves—that is, generate a stream of financial returns whose present value exceeds the cost of the projects. On solvency grounds, deficit financing of those projects—termed "self-liquidating" (Mintz and Smart 2007)—is most likely to be justified, because such projects increase government net worth, even if they raise public debt in the short run. In practice, for the projects to increase government net worth,
the government must be able to capture the returns. For the first type of projects, user fees must be sufficient to cover project costs; for the second type of projects, taxes must be high enough to translate the additional growth into sufficient additional revenues.

In the absence of user fees, many growth-enhancing projects may fail to generate sufficient tax revenues to cover their cost. With the low (marginal) tax collection rates of many developing economies, the growth impacts have to be considerable to yield the required tax revenue increases. For example, with a tax rate of 0.2, the output contribution would have to be five times as high as the project’s user cost for the government to break even. So if the user cost is about 10 percent (say, a 5 percent real interest rate and a 5 percent rate of depreciation), the project’s marginal productivity must be at least 50 percent for the given tax rate to yield sufficient revenues (see Servén 2007 for the analytics of this and similar calculations).

Such high productivity is more likely to arise in situations in which the initial endowment of public capital is low (relative to that of other productive assets)—specifically, when public capital services are substantially underprovided so that the marginal product of capital exceeds its user cost by a wide margin. Empirically, the international evidence appears to be consistent with the view that the marginal productivity of infrastructure capital is higher in developing economies, especially poorer ones, than in industrial countries (Calderón and Servén 2007).

Probably as a reflection of these country-specific ingredients, empirical results are mixed regarding whether public investment may be self-financing through its growth and tax-collection effects. Perotti (2004) examines this issue in five countries in the Organisation for Economic Co-operation and Development (OECD), using a vector autoregression approach. He finds that in Canada and the United Kingdom, the extra public capital makes a negative growth contribution: in Australia and the United States, the growth and tax collection effects finance only 20–30 percent of the investment cost. Only in Germany do these effects finance more than 100 percent of the cost. Using similar techniques but a broader sample of industrial countries, Pereira and Pinho (2006) find that public investment is roughly self-financing in France, Greece, and Ireland and more than self-financing in Germany and Italy. The growth effects are large in the majority of countries considered.

Because developing economies possess smaller infrastructure capital endowments than industrial countries do, infrastructure capital might be expected to have a higher marginal productivity and so to come closer to being self-financing. Ferreira and Araujo (2005) find that public infrastructure investment is self-financing in Brazil, although it takes 10 years or more for the government to collect sufficient tax revenues to recoup the investment cost.
Even if public sector projects fall in the intermediate area of having high returns for the economy as a whole but insufficient returns for public finances to improve public sector solvency, it may still be suboptimal to cut such projects during periods of fiscal austerity. An ideal marginal revenue collection scheme would allow the public sector to capture the returns, thereby eliminating the wedge between economywide and public sector returns. After all, the business of the public sector is precisely to provide public goods that yield a high return for the economy as a whole.

In addition, public investment projects are more likely to exhibit higher marginal productivity ex post if the government's ex ante project evaluation capabilities are sufficiently strong that they select high-return projects and reject low-return ones. This, however, is far from assured in practice. Many developing economies lack ex ante and even ex post project evaluation capabilities. (One exception is Chile, which has thorough procedures for evaluating projects: Fontaine 1997.)

Unconditional endorsement of public infrastructure spending would lead to wasteful investments, as Tanzi and Davoodi (2002) convincingly argue. Roads would be built to nowhere, and useful roads would not be maintained. Power plants would lie idle after being built too far ahead of demand. Water supply networks would be fully used but still burden the budget, because the tariff increases on which their financial viability was predicated were not allowed. Some investments would be well motivated but poorly informed; others would be motivated by bribes, patronage, or photo opportunities (Kefer and Knack 2007).

Infrastructure investment is the expenditure item that has attracted most attention in the ongoing debate about the design of fiscal policy. But the link between spending composition and solvency arises in a broader context. On the one hand, not all public investment projects yield future income to the government. On the other hand, some current expenditures do yield future fiscal returns. Infrastructure operation and maintenance expenditure is a case in point. Operation and maintenance determines the useful life of capital and hence has a "capital-creating" effect similar to that of investment. If public capital yields financial returns to the government, so does operation and maintenance. In fact, the financial, as well as social, return on operation and maintenance expenditure may well exceed that of new capital when the assets are not being properly maintained (Rioja 2003a, b; Kalaitzidakis and Kalyvitis 2004; Servén 2007).

This does not imply that developing countries should rush to raise public investment or that, as a rule, public investment increases should be financed with debt (or in any other particular way). The decision to invest should be guided by the return on the investment. That return is determined primarily by the marginal productivity of public capital, itself dependent on the government's ability to select good projects and the (relative) scarcity of services rendered by public capital (for example, the availability of infrastructure services). Both return and
cost calculations should embody risk adjustments to take account of uncertainty. Indeed, assessments of the effect on net worth of public investment projects should err on the side of caution, particularly when the government’s initial indebtedness is high, because in such cases even small changes in interest rates may have very large adverse effects on public finances.

Excluding Certain Public Investments from Fiscal Targets

Broadly speaking, there are two possible ways to address the bias against productive public spending implicit in existing fiscal targets. One is to retain the targets but exempt from their action certain public investments deemed more likely to enhance growth and solvency. The other is to adopt new fiscal targets. This section reviews the first alternative; the next section discusses the second.

Privately Financing Public Investment Projects

One way to place investment projects beyond the reach of short-term deficit and debt targets is to have private firms finance them. Indeed, across the developing world, many governments have turned to the private sector to finance new investments. In Latin America, for example, many governments have privatized their telecommunications firms and parts of their power and water industries. In the transport sector, private firms are often engaged in public–private partnerships in which the government retains an important financial role but the private sector finances investment. Chile and Colombia have had roads privately financed under arrangements in which the government provides revenue or foreign exchange guarantees. Other countries have begun to use a different form of public–private partnership in which a private firm finances an asset (such as a school, hospital, or prison), but the government purchases the service under a long-term contract.

These arrangements may improve the returns to investment and thus enhance government solvency. In many cases, however, concerns about efficiency and solvency have played a minor role, and the resort to private financing has been guided primarily by the desire to evade the pressure of liquidity targets on public investment. Projects conceived with such a purpose in mind may not be well designed from the point of view of efficiency or solvency.

The difference between liquidity and solvency effects is particularly apparent in privately financed projects in which the government purchases the service under a long-term contract. In such projects, explicit debt is replaced by similar commitments that are typically off balance sheet, without any major change in the magnitude of the government’s financial obligations. It is also apparent when the government provides guarantees to private investors—such as guarantees of the
private firm's debt or revenue—that leave the public sector bearing much of the investment risk (Hemming and International Monetary Fund 2006; Irwin 2007b). Even when such guarantees are not formally offered ex ante, they may be provided ex post through renegotiation of concession agreements. The bailout of the Mexican toll road program in 1997, for example, cost 1.0–1.7 percent of GDP (World Bank 2005; see also Guasch 2004).

On the whole, private financing has not come to play the dominant role in the provision of infrastructure services in Latin America or elsewhere that some observers expected. Although private financing now dominates telecommunications and some other infrastructure industries in some countries, it still plays a small role in roads and water and sanitation—something that is unlikely to change in the near term. Moreover, it would be undesirable for decisions about the ownership of infrastructure firms to be driven by short-term fiscal constraints. When private ownership works better than public ownership in terms of efficiency, equity, or both, a state-owned firm should be privatized, even if ownership of the firm requires little investment. Likewise, when public ownership works better, the firm should remain public, even if major investment is required. Private financing may thus lessen the problem caused by liquidity targets in certain cases, but it is not an appropriate response to the general problem.

**Excluding Specific Public Projects**

Another option is to exclude from fiscal targets certain investments undertaken by the public sector. A recent proposal would exclude projects financed by multilateral institutions on the grounds that such projects are more likely than others to be carefully screened and designed. This idea has not garnered much support, partly because the fungibility of money means that the marginal financing from multilateral institutions would not necessarily support the intended projects. Furthermore, in many developing economies, total multilateral flows are too small to make a big difference.

A second proposal, developed and refined by the International Monetary Fund (IMF 2004), is to exclude from fiscal targets investments by public enterprises that are deemed to be commercially run. This proposal (which is not new; see Afonso 2005) is, in principle, potentially important for countries in which public enterprises are included in the public sector aggregates monitored under fiscal programs. This is the case in Latin America but not in most other regions.

In practice, this approach poses several problems. First, appropriate criteria to identify commercially oriented public enterprises are difficult to establish. Second, the fact that enterprises that meet likely criteria are the exception rather than the rule in many countries (IMF 2005) detracts from the practical relevance of this approach for public investment.
Another difficulty is that excluding commercially run public enterprises from targets may further restrict investment elsewhere in the public sector if those enterprises make a positive net contribution to the aggregate budget surplus. Exclusion of those enterprises would then make fiscal targets more, rather than less, stringent for productive public expenditure. One way around this difficulty is to exclude the investments, but not the savings, of these firms from fiscal targets. Doing so, however, adds complexity, which detracts from the transparency of the approach. An alternative would be to relax the targets by the net saving of the excluded public enterprises.

The fundamental problem with this proposal is that investments by enterprises that are not commercially run may still have offsetting fiscal and economic benefits—as is the case of many investments in roads, for example. In other words, the investments of commercially run public enterprises may not be the ones with highest priority from a fiscal or social perspective. Removing restrictions only on investment by commercially run public enterprises may still leave an overall public investment program far removed from the socially desirable one. For example, it is unlikely that Brazil's infrastructure needs would be significantly alleviated by allowing more investment by the commercially run public oil company PETROBRAS. Such outcomes are a general problem with any proposals on the basis of the special treatment of specific investments.

Developing New Fiscal Targets Incorporating Measures of Net Worth

The limitations of "selective" approaches that exempt certain investments suggest that it is important to consider more fundamental changes, in particular, whether governments can develop measures of net worth that are sufficiently accurate and objective to be used as a basis for fiscal targets. Measuring net worth—the difference between the value of assets and the value of liabilities—requires forecasts. The value of an asset is the present value of the net revenues it will generate and such revenues that are usually uncertain. Likewise, the value of a liability is the present value of the payments it will cause the government to make, and such payments are often uncertain. Because of these uncertainties, indicators of net worth are inherently approximate.

Governments seeking to beautify their reported fiscal positions can take advantage of uncertainty to overestimate future revenues and underestimate future costs. Governments seeking to protect some category of public spending on political grounds can exaggerate the value of the "assets" it creates, even if the spending is actually wasteful. If an indicator of net worth is too vulnerable to such manipulation it has too little credibility to be useful.
Two indicators are used to measure net worth, one generated by modern accrual accounting, the other by long-term fiscal projections. This section examines the accuracy and reliability of each.

**Modern Accrual Accounting**

Like traditional cash accounts, modern accrual accounts include information on short-term cash flows. Unlike traditional cash accounts, accrual accounts also include a balance sheet showing assets and liabilities. As a result, accrual accounts generate a measure of net worth. They also include a measure of the surplus or deficit that is not based on current cash flows. That measure includes revenues that have been earned but are not yet collected and bills that are payable but not yet paid. Crucially, investment itself is not counted as an expense in the period of investment; only the depreciation of the investment is included. The difference between accrual revenues and accrual expenses gives an income-statement surplus that is roughly equal to the increase in the government’s net worth.

To see the implications of accrual accounting, consider a government investing $200 million in a power plant, financed entirely by borrowing (this example is taken from Irwin 2007a). Assume that, in the first year, no revenue is received, no operating costs are incurred, and no depreciation occurs. Under cash accounting, the government’s accounts show $200 million in extra expenditure, which increases the cash deficit and debt by the same amount (table 1). Under modern accrual accounting (table 2), these consequences of the investment on the government’s cash flows and debt are revealed, but so, too, are the consequences for the government’s assets. The accounts report that the investment has no net effect on the government’s net worth or income-statement surplus.

Accrual-based accounting standards for financial reporting have the advantage of being designed to limit the self-serving bias that uncertainty makes possible. Most obviously, financial reports, whether based on cash or accrual accounting,

<table>
<thead>
<tr>
<th>Table 1. Debt-Financed Investment in Cash Accounting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
</tr>
<tr>
<td>Revenues</td>
</tr>
<tr>
<td>Expenditure</td>
</tr>
<tr>
<td>Surplus</td>
</tr>
<tr>
<td>Debt</td>
</tr>
</tbody>
</table>

*Note: Cash surplus is the sum of cash disbursed to operations and cash disbursed to investment. Source: Authors.*
must be audited by an independent auditor. In addition, accrual accounting standards tackle bias by preferring measures that are objectively verifiable even at the cost of some relevance. Some standards, for example, require an asset to be valued by recording its acquisition cost and then depreciating the cost according to a simple formula. The resulting value can only approximate the asset’s true value, but the measure is less vulnerable to bias than alternative measures. When standards require the reporting of market value, they sometimes require the valuation to be performed by an independent expert (other than the auditor). Accounting scandals show that these safeguards can fail to prevent biased reporting, but they are surely better than nothing. Moreover, although the risk of misleading accrual information is real, it does not provide a strong argument against the adoption of accrual accounting, because cash-based reporting is at least as vulnerable to manipulation, as argued earlier.

Reporting according to modern accrual accounting standards is, however, more costly than reporting according to cash standards, and it can take years for a government to move from cash to accrual accounting. These costs are more likely to be justified in middle- and high-income countries than in the poorest developing economies. Many high-income countries (including Australia, Canada, New Zealand, the United Kingdom, and the United States) have already made the transition, and many middle-income countries (including Chile, Indonesia, the Philippines, and South Africa) are adopting accrual accounting.

Although accrual accounting generates valuable information missed by traditional cash accounting, it is not sufficient for the assessment of net worth, even

---

**Table 2. Debt-Financed Investment in Modern Accrual Accounting**

<table>
<thead>
<tr>
<th>Item</th>
<th>Amount ($ million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income statement</td>
<td></td>
</tr>
<tr>
<td>Revenue</td>
<td>0</td>
</tr>
<tr>
<td>Expenses</td>
<td>0</td>
</tr>
<tr>
<td>Income-statement surplus</td>
<td>0</td>
</tr>
<tr>
<td>Balance sheet</td>
<td></td>
</tr>
<tr>
<td>Assets</td>
<td>200</td>
</tr>
<tr>
<td>Liabilities</td>
<td>200</td>
</tr>
<tr>
<td>Net worth</td>
<td>0</td>
</tr>
<tr>
<td>Cash-flow statement</td>
<td></td>
</tr>
<tr>
<td>Cash disbursed to investment</td>
<td>200</td>
</tr>
<tr>
<td>Cash surplus</td>
<td>-200</td>
</tr>
<tr>
<td>Cash from financing</td>
<td>200</td>
</tr>
</tbody>
</table>

*Note:* Cash surplus is the sum of cash disbursed to operations and cash disbursed to investment.

*Source:* Authors.
in middle- and high-income countries. For one thing, accounting values can diverge too much from true values. After many years of inflation, the depreciated acquisition cost of an asset may greatly underestimate the present value of the cash flows it will generate. In addition, accounting values of assets that generate user fees and higher tax payments at best capture only the value of the user fees, because the present value of future tax revenues does not count as an asset from the conventional accounting perspective. In contrast, durables are generally treated as assets and valued at their depreciated acquisition or replacement cost, even if they generate no future cash flows from either user fees or taxes. Expenditure on a bridge to nowhere can create an accounting asset even if it generates no tolls and does nothing to increase economic output.

**Long-Term Fiscal Projections**

Long-term fiscal projections have the potential to remedy some of the shortcomings of accrual accounting. Such projections, prepared in various ways by countries such as Australia, New Zealand, the United Kingdom, and the United States, can include estimates of the government's operating and investing cash flows over the next 50–75 years, which can then be discounted back to the present to arrive at an estimate of the government's net worth. Crucially, all expected cash flows under current policies can be projected, including taxes and welfare expenditure.

The projections can include public investment, expenditure on operations and maintenance, and payments to privately financed firms in public-private partnerships. Revenues from user fees can be included. If evidence suggests that some investments increase tax revenue by generating more taxable economic activity, this extra revenue can also be included.

Long-term fiscal projections can be designed to take account of the uncertainty of future cash flows. On the one hand, future revenues and expenditure can be adjusted for risk: risky tax revenues, for example, can be discounted at a higher rate than more predictable pension spending. On the other hand, the projections can show how net worth changes with critical assumptions about life expectancy, health care costs, and output. If output is modeled as a function of the stock of public capital, the projections can show how sensitive the government's net worth is to this assumption.

The biggest disadvantage of long-term projections is a corollary of their usefulness: generating the relevant information requires estimates that are subject to enormous uncertainty. What will the future rate of growth of GDP be? How will it be affected, if at all, by public investment? The large extent of reasonable disagreement about such estimates implies a large range of reasonable estimates of the government's net worth. This makes room for self-serving projections.
Table 3. Benefits and Drawbacks of Alternative Sources of Fiscal Indicators

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cash accounting</th>
<th>Modern accrual accounting</th>
<th>Long-term projections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides information on short-term cash flows</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Provides information on net worth, given current policies</td>
<td>No</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>Incorporates uncertainty</td>
<td>Avoids the issue</td>
<td>Partially</td>
<td>Yes</td>
</tr>
<tr>
<td>Limits self-serving forecast bias</td>
<td>Yes</td>
<td>Yes</td>
<td>Not easily</td>
</tr>
</tbody>
</table>

Source: Authors.

Though uncertainty in the estimates is unavoidable, a government can make its projections more credible. It can allow others to see how the results are generated by making its projection model publicly available (in a spreadsheet on the Internet, for example). For critical parameters or key variables, it can use the estimates of a panel of independent experts, as Chile does in estimating its structural surplus. It can prepare and publish standards that it will follow in making the projections. And it can legislate that an independent auditor must opine on whether the projections follow those standards and reflect the stated assumptions. International organizations and financial institutions could help develop the standards and expertise necessary for some of these steps.

Each approach has benefits and drawbacks (table 3). Cash accounting offers the necessary information on liquidity and, because of its short-term focus, limits opportunities for some sorts of bias, but it provides no information on the crucial issue of government net worth. Modern accrual accounting fills this gap using methods designed to limit bias, but partly because of the concern to limit bias, it includes some poor estimates of values and provides no information on crucial elements such as future tax revenues. Long-term projections can overcome these problems, but only at the cost of requiring more estimates, creating more leeway for bias. Given the advantages and disadvantages of each approach, the best strategy, at least for middle- and high-income countries, would seem to be for governments to develop both accrual accounting and long-term projections while continuing to monitor short-term cash flows.

Fiscal Targets and Fiscal Rules

Indicators of net worth can also form the basis for new fiscal targets and rules that promote solvency without sacrificing public investment. One alternative is the so-called golden rule, according to which governments can borrow only to
invest in the creation of new assets. This idea of separating the current and capital budgets is hardly new, but it has recently been revived (a classic reference is Musgrave 1939; see Bassetto and Sargent 2005 for historical background). The British government has adopted a version of the golden rule that "over the economic cycle" allows it to borrow only to invest and only as long as public sector net debt remains below 40 percent of GDP (H.M. Treasury 2004).

Accrual accounting suits the golden rule, because another way of stating the rule is to say that governments must not run an income-statement deficit (that is, they must not reduce their net worth). Because the income-statement surplus excludes expenditure on investment, the golden rule encourages investment. Because the income-statement surplus includes depreciation, however, adopting the golden rule is not the same as simply exempting investment from the fiscal targets—it amounts to exempting net investment from fiscal targets. Observance of the golden rule over a long period of time would eventually result in a public debt stock no larger than the public capital stock so that to a first approximation the outstanding debt would be fully backed by public assets.

The limitations of typical accounting mean that the golden rule does not ensure solvency or even expected solvency. The reason is that the assets may not yield an expected return high enough to cover the interest on the debt that financed their acquisition. Furthermore, by treating current and capital spending differently, the golden rule offers an incentive for opportunistic misclassification of expenditures.

An alternative that, in principle, avoids these problems is the permanent-balance rule. Roughly stated, it requires governments to set tax rates at a constant fraction of output that over the long run pays for the government's present and future expenditure (Buitter and Grafe 2004). Named by analogy with Milton Friedman's permanent-income hypothesis, the rule allows governments to borrow when revenue is temporarily low or when current investment opportunities are greater than future investment opportunities. Long-term fiscal projections suit the permanent-balance rule, because such projections are required to determine the required minimum tax rate. Implementing the permanent-balance rule successfully would thus require addressing the problems discussed earlier about the reliability of long-term fiscal projections, which is no easy matter.

A compromise solution that avoids some of the problems with both the golden rule and the permanent-balance rule is the modified golden rule recently proposed by Mintz and Smart (2007). They suggest that governments should be able to borrow to invest in self-liquidating assets, such as power plants that generate future revenues for the government but not in assets that generate no revenues, such as typical public schools. Furthermore, just as firms do not finance all their assets with debt, governments should be limited to borrowing only some fraction of the value of the revenue-generating assets. These features would result in a rule that is
more conducive to solvency than the golden rule, without relying as much on potentially unreliable long-term forecasts as does the permanent-balance rule.

Conclusions

In many industrial and developing economies, governments have cut back on public investment as they have brought their budgets closer to balance. Although budget cuts were probably necessary, the cuts in public investment may have been counterproductive, because much theory and evidence suggest that public investment has the potential to increase future output.

In the worst case, investment cuts trigger a vicious circle, in which the subsequent deterioration of future revenue forces further investment cuts, leading to yet further deterioration, further investment cuts, and so on. What is supposed to be fiscal adjustment in this case actually has the same consequences as fiscal profligacy. Cutting investment to promote solvency becomes the fiscal equivalent of walking up the down escalator—riders step up only to end up below where they started.

The cuts in public investment should have come as no surprise when most countries measure their fiscal position not in terms of net worth but in terms of short-term cash flows and gross debt, and cutting investment can reduce debt and short-term cash flows, even as it reduces net worth. The problem afflicts both industrial and developing economies, but it is more pressing in developing economies, which have not yet built up their public capital stocks.

The decline in public investment suggests the need to rethink fiscal strategies. In some cases, it may be best to increase public investment and accept a higher short-term cash deficit in exchange for higher tax and user-fee revenues later. This strategy is unlikely to be right for all countries, however. Those with good infrastructure and bad fiscal positions may indeed do well to cut public investment. Countries with high taxes and debt may do best to increase public investment but finance it by cutting current expenditure. Still others, with high debt and little room for cuts in current expenditure, may have no choice but to raise taxes or forgo improvements in their infrastructure. Each case must be analyzed on its merits, with—given the tendency to be optimistic in forecasting growth and the performance of investments—a degree of skepticism. One general lesson is that appropriate spending composition has to be an essential part of fiscal adjustment and consolidation strategies, because it affects growth outcomes. In other words, spending targets and growth forecasts cannot be set without regard to the composition of expenditure, as they currently are.

All governments are likely to benefit from better fiscal information. The idea is not to abandon measures of debt and short-term cash flows, which are clearly
important, but rather to supplement them with measures of assets, yielding a
measure of net worth and its change over time. What is needed is information
that allows governments to quickly determine when improvements in short-term
cash flows are coming at the expense of declining net worth. Two means of gener-
ating such information are constructing long-term projections of fiscal cash flows
and adopting modern accrual accounting.

Better fiscal information is helpful irrespective of whether the government follows
quantitative fiscal rules or targets. But the question also arises whether govern-
ments should set themselves fiscal rules or targets incorporating measures of invest-
ment or net worth. Because debt and short-term cash flows matter, rules or targets
based exclusively on net worth may not be helpful. But combining net worth with
conventional fiscal measures may have merit. The United Kingdom, which adopted
a version of the golden rule combined with a debt target, offers one example. An
even better option might be a modified golden rule that allows borrowing to
finance a portion of the cost of cash-generating assets but also requires that some
proportion be financed by current taxes. There is no obvious “best” solution, but
whatever the specific solution chosen, it is clearly time to change the exclusive
focus on public sector liabilities and bring public sector assets into the picture when
designing fiscal adjustment. It is much easier to walk up the up escalator.

Notes

William Easterly is professor of economics at New York University, where he is also a faculty affiliate
of Africa House and co-director of the Development Research Institute; his email address is william.
easterly@nyu.edu. Timothy Irwin is a senior economist in the Finance, Economics, and Urban
Department at the World Bank; his email address is tirwin@worldbank.org. Luis Servén (corre-
responding author) is research manager for macroeconomics and growth in the Development
Research Group at the World Bank; his email address is lserven@worldbank.org. An earlier version
of this article was prepared for the World Bank’s Latin American Regional Studies Program. The
authors are grateful to Penelope Brook, Antonio Estache, José Luis Irigoyen, Guillermo Perry, Sergio
Rebelo, Augusto de la Torre, and three anonymous referees for helpful comments.

References

Afonso, J. 2005. “Fiscal Space and Public Sector Investment in Infrastructure.” IPEA Texto para
Discussao 1141, Instituto de Pesquisa Econômica Aplicada, Brasilia.


Easterly, Irwin, and Servén


What Can Countries in Other Regions Learn from Social Security Reform in Latin America?

Indermit S. Gill, Ceren Ozer, and Radu Tatucu

About a dozen countries in Latin America have enacted reforms that include elements being contemplated elsewhere, including the partial privatization of social security. It is not easy to draw universal lessons for social security reform from the experience of countries such as Argentina, Chile, and Mexico, however, where sizeable public pension systems went bankrupt before the populations aged, mainly because of mismanagement. Most developing economies have much smaller social security systems. Relatively well-managed systems in industrial countries face problems that are long term in nature and have been brought about by an aging population. The experiences of Latin America nevertheless offer some general lessons for countries in other parts of the world. These lessons relate to changes in labor market incentives accompanying reforms and how workers react to them, government actions that have met with success in managing the transition to funded pensions, and the expectations of individuals from social security systems. Latin America’s reforms suggest that the most effective approach is to keep payroll taxes low, governments solvent, and social security systems focused on providing reasonable insurance against poverty in old age. JEL codes: G23, H31, H53, H55, J26.

Latin America has long experience with social security reforms. Since the early 1980s, especially during the 1990s, about a dozen countries in the region have attempted to radically reform their social security systems. Many observers have studied social security reforms and outcomes in Latin America to draw lessons. Among the more widely discussed social security reform experiences have been those of Chile, Argentina, and Mexico. Gill, Packard, and Yermo (2005), among others, have analyzed these experiences, with the objective of informing pension
Table 1. Actual and Projected Total Fertility and Life Expectancy at Birth, by Region

<table>
<thead>
<tr>
<th>Region</th>
<th>1970–75</th>
<th>2000–05</th>
<th>2045–50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>fertility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>World</td>
<td>4.49</td>
<td>2.65</td>
<td>1.56</td>
</tr>
<tr>
<td>Africa</td>
<td>6.72</td>
<td>4.97</td>
<td>2.03</td>
</tr>
<tr>
<td>Asia</td>
<td>5.08</td>
<td>2.47</td>
<td>1.42</td>
</tr>
<tr>
<td>Europe</td>
<td>2.16</td>
<td>1.40</td>
<td>1.33</td>
</tr>
<tr>
<td>Latin America and the Caribbean</td>
<td>5.05</td>
<td>2.55</td>
<td>1.36</td>
</tr>
<tr>
<td>Northern America</td>
<td>2.01</td>
<td>1.99</td>
<td>1.35</td>
</tr>
<tr>
<td>Oceania</td>
<td>3.23</td>
<td>2.32</td>
<td>1.42</td>
</tr>
</tbody>
</table>


policy in Latin America. This article attempts to draw lessons for other countries from the Latin American experience.

As in Latin America, many other developing regions are experiencing significant demographic change. Birth rates have fallen, and life expectancy is on the rise (table 1).

Latin America offers the most varied experience with structural pension reform. Thus it provides insights into how radical reform of pension systems can help meet the pressures created by rising longevity and persistent old-age poverty.

The fiscal deficits and increasing contingent liabilities (obligations to pay sums dependent on future events) of generous public pension systems, often combined with system mismanagement, created an immediate impetus for governments to institute structural pension reforms in Latin America. Rising pension costs raise questions of fiscal sustainability across the globe, but so far only a few countries have engaged in major structural reform. Many focused instead on parametric changes—adjusting the size and scope of their single-pillar social security systems (the mandatory, pay-as-you-go, publicly provided part) by changing the rates of contributions, the benefit calculations, and the retirement age. This article assesses whether the demographic and social pressures these countries face require structural reform of social security, along the lines of the reforms adopted in Latin America.

This article does not address the special challenges of reforming pension systems for government workers: it focuses on the social security system for the employees of enterprises. Civil service pensions put significant fiscal pressures on more than half of the world’s countries—including some of the largest developing economies, such as Brazil, China, and India—which have separate pension schemes for civil servants. Civil service pension reform is a contentious issue. Not
surprisingly, in most Latin American countries the reforms did not affect private and public sector employees equally. For political reasons reforming governments often avoided structural changes to the public pension systems benefiting the military and civil servants.

The article is organized as follows. The next section summarizes social security problems and reforms in various regions. The following section assesses the performance of reform in Latin America and draws lessons from that experience for other countries. The last section draws implications for policymakers considering reforms.

Social Security Problems and Reforms

The underlying conceptual framework is based on the “comprehensive insurance” concept of Ehrlich and Becker (1972, 2000). In the face of a possible loss a “comprehensive insurance” approach suggests that individuals can insure against the loss, take steps to lower the likelihood that the loss will occur, or do nothing. The purchase of insurance transfers income from “good” to “bad” times in order to reduce the magnitude of losses in bad times. Individuals can insure themselves in two ways: through mechanisms that pool the risk of the loss occurring among those who are exposed to this risk or by consumption smoothing through individual savings (“self-insurance”). In a perfect world, pensions could be left to private insurance and to individuals’ voluntary saving decisions. In the presence of imperfect information, missing markets, and other distortions government involvement becomes necessary.

In addition to insurance and consumption-smoothing objectives the two other primary objectives of pensions are poverty relief and redistribution. Redistribution complements the role of progressive taxation, for example, by subsidizing the consumption smoothing of individuals who earned little during their working years (Barr and Diamond 2006). Pension policy may also have secondary goals, such as improving the operation of labor and capital markets and encouraging individuals to save more. Some might argue that promoting economic growth is an additional objective of pensions. One of the key debates on pensions centers on the relative weights of these different objectives.

Another debate concerns whether pensions should be pay-as-you-go or funded. Most state-run pension schemes remain pay-as-you-go. Private schemes are generally funded, with pensions paid from a fund built up over time by members’ contributions. Many countries in Latin America added mandatory contributions to private-funded pensions to their existing pay-as-you-go schemes.

There is general agreement on the desirability of regulated voluntary pensions; disagreements remain over whether private schemes should be mandatory.
Supporters of the mandatory-funded private schemes emphasize that individual incentives to work, save, and incur risks are least distorted if the mandated saving flows into privately managed accounts. Individuals' pension-related decisions involve long-run choices and require a good understanding of pensions products, which are often complex. Both factors create information problems, which reduce—often considerably—people's ability to make choices that maximize their long-term well-being (Barr and Diamond 2006). Arenas and Mesa-Lago (2006) find, for example, that many workers in Chile lack the data and skills to make an informed selection of the best pension provider. Barr and Diamond (2006) argue that imperfect information in this context cannot be addressed simply by offering more information, because the issue at hand is an information-processing problem.

The most common component of a pension system is a national defined-benefit scheme in which pension benefits depend on a worker's wages and age (table 2). Some countries also require that workers contribute to individual retirement accounts, and pensions are paid from the accumulated funds. Many countries also encourage workers to contribute to individual accounts managed by financial institutions. The last two kinds of pensions are called defined-contribution plans (for a more detailed explanation of the economics of pensions, see Barr and Diamond 2006).

A recent review of pension systems in 53 countries—all 30 Organization for Economic Co-operation and Development (OECD) members plus 23 countries in Eastern Europe and Central Asia, Latin America and the Caribbean, and the Middle East and North Africa—shows that the primary differences across systems lie in three main areas (Whitehouse 2007). First, the level of social security

Table 2. Instruments for Old-Age Income Security

<table>
<thead>
<tr>
<th>Nature of instrument</th>
<th>Main function</th>
<th>Main role of government</th>
<th>Principal risk bearer</th>
<th>Financial instrument</th>
<th>Mainstage: pooling</th>
<th>Mainstage: saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common name</td>
<td>First pillar</td>
<td>Deline benefits</td>
<td>Government</td>
<td>Unfunded pay-as-you-go</td>
<td>Mandatory</td>
<td>Voluntary</td>
</tr>
<tr>
<td>Main function</td>
<td>Insure against poverty in old age, reduce income inequality</td>
<td></td>
<td></td>
<td></td>
<td>Second pillar</td>
<td>Third pillar</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Smooth consumption over life cycle</td>
<td>Smooth consumption over life cycle</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source: Authors’ compilation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

benefits (the target replacement rates) varies considerably. There is a negative link between the target replacement rate for mandatory pensions and the relative importance of voluntary private pensions. In countries such as Canada, the United Kingdom, and the United States, where voluntary private pension provisions are widespread, mandatory pensions are relatively small.

Second, the relative emphasis on pooling and saving differs significantly. Countries such as Australia, Canada, New Zealand, and the United Kingdom emphasize the pooling function; Nordic countries and most Latin American countries also have progressive systems. In continental Europe (outside the Nordic countries) and the Middle East and North Africa pensions are more strongly correlated with earnings, signaling a greater emphasis on the savings function.

Third, the relative reliance on public and private sectors varies greatly. In addition to the larger role for voluntary and private provision in countries with low target replacement rates, many countries also involve the private sector in running the mandatory pension system. The best-known cases of pension reforms that increase the role of the private sector are in Latin America and the Caribbean, followed by Eastern Europe and Central Asia. The private sector plays an important role in mandatory pension provision in about one-third of the high-income OECD countries. At the other end of the spectrum, countries in the Middle East and North Africa do not yet involve the private sector in mandatory pensions, and voluntary provisions barely exist.

The experiences of Brazil, China, India, the Russian Federation, and South Africa illustrate the range of problems faced by developing economies and transition economies. Reforms in these and other countries are described below.

**Brazil**

Conditions in Brazil today resemble those in some Latin American countries before they undertook the type of structural reforms assessed below. Pension reform was motivated mainly by fiscal pressures. Despite having a young population, Brazil’s level of public expenditure on pensions is large. Subsidies to cover Brazil’s public pension regimes’ deficits rose from 4.6 percent of GDP in 1998 to 5.6 percent in 2004 (Giambiagi and de Mello 2006). So far pension reform has involved streamlining the system’s mandatory first pillar and developing a third pillar of voluntary, complementary, personal saving schemes, without creating a second pillar of mandatory individual saving plans. Since the late 1990s, reform of the regime for private sector workers has been aimed at tightening eligibility conditions, reducing replacement rates, and increasing the share of population covered by social security.
East Asia

East Asia has a diverse set of pension systems. Malaysia and Singapore have large provident fund systems operating under public administration at the national level on the basis of defined contributions. The Republic of Korea, the Philippines, and Thailand have OECD—style defined-benefit pension schemes (though with lower coverage rates than industrial countries), with more emphasis on redistribution.

Like other communist countries, China formerly had defined-benefit pay-as-you-go pension systems covering urban public sector employees, and contributions were largely the responsibility of state-owned enterprises. Rising pension expenditures (caused partly by the use of early retirement as a mechanism to deal with excess workers at state enterprises) and declining contributions (the result of poor performance by state enterprises, rising unemployment, a growing informal sector, and weak enforcement) motivated China to seek the best reform alternatives (Asher and others 2005).

South Asia

India, like other countries in South Asia, is still at the beginning of its demographic transition. It has a low ratio of pensioners to workers, who continue to contribute to social security schemes. Although only 13 percent of India’s labor force is covered by pensions, pension debt is becoming a serious issue. Implicit pension debt is estimated at 25 percent of GDP nationally, and in some states the extent of the problem is much greater. India is in the process of passing into law a new pension system that would shift all new central government employees to a defined contribution plan from the current defined-benefit scheme, shifting the risk of retirement financing from the government to individuals (Shah 2006). Participants in the new scheme will have access to a range of investment products from selected private sector companies. The new pension system will be offered on a voluntary basis to private sector workers. Aiming to fulfill the social protection dimension of pensions, India’s noncontributory pensions target the elderly poor: the means-tested schemes administered by states and supplemented with federal funds reach 1 of every 10 elderly Indians.

Eastern Europe and Central Asia

Regionally, Europe and Central Asia is second only to Latin America in terms of pension reform activity. In the early transition period, countries in the region faced serious challenges to their social security systems as output fell, contributions declined, and the number of beneficiaries grew. Transition proved to be
challenging politically. Some countries chose parametric reforms, such as raising the retirement age, with or without changing the benefit formulas; others, especially the European Union accession countries, undertook structural reform. Ten countries in Europe and Central Asia created second pillars of mandatory, funded individual accounts.

The Russian Federation began structural reforms of its pension system in 2001. The new system comprises three pillars. The first pillar, the major component of the system, is a publicly managed pay-as-you-go, defined-benefit scheme that consists of a flat basic benefit and a notional defined-contribution scheme. The second pillar is a mandatory defined-contribution scheme with mixed public–private management. The third pillar is the voluntary privately managed component (OECD 2006).

Middle East and North Africa

Countries in the Middle East and North Africa have already put in place defined-benefit pension systems financed on a pay-as-you-go basis. Egypt, Iran, and Libya have also developed noncontributory pension schemes. Even with young populations, many pension systems in the region are not financially sustainable without reform. Despite this, reforms have been limited, with Lebanon and Morocco among the few countries considering systemic pension reform.

Sub-Saharan Africa

Noncontributory schemes in Sub-Saharan Africa exist in only a few countries, such as South Africa, where they are financed by general revenues. As in South Asia, the pension reform agenda in Sub-Saharan Africa is driven by the fiscal pressures arising from civil service pensions. Average coverage in the region is less than one-fifth of the labor force, with the rest of the population relying on its own resources and informal old-age support.

Pension systems around the world are thus diverse, with every country facing a unique set of problems. But there are some common features. Many countries in the developing world—including major economies such as China and India—have to deal with expanding social security systems that cover only small parts of their populations, and they have to do so within tight fiscal constraints. Industrial countries face problems that are long term in nature and have been brought about by an aging population. While the immediate concern behind the reform process has often been fiscal sustainability, getting the incentives right is equally important.
Performance in Latin America and Lessons of Experience

The experiences of Latin America offer some general lessons for countries in other parts of the world. Chile first adopted structural reform in 1981. Argentina, Bolivia, Colombia, Mexico, Peru, and Uruguay followed in the 1990s, and Costa Rica, El Salvador, the Dominican Republic, Ecuador, and Nicaragua followed after 2000. The notable exceptions to reform are Brazil and República Bolivariana de Venezuela.

The details of the reforms vary across countries. What is common is that a publicly mandated and administered pay-as-you-go component operated on a defined-benefit basis was retained and a publicly mandated but privately administered system of defined-contribution personal accounts was added. Governments also made some attempts to increase voluntary saving through defined incentives, such as Investment Retirement Accounts in the United States, which encourage individual retirement savings through tax benefits.

But the headline item has been the new system of personal accounts. Mandatory personal accounts constitute the second pillar. Second-pillar reforms can be classified into three categories. The first is the “Chilean model,” which made private accounts mandatory for all new workers; Bolivia, El Salvador, and Mexico also adopted this model. The second is what might be called the “Peruvian model,” which Colombia also adopted. Under this model, new workers are given a choice between a downsized pay-as-you-go pension and a private account. Under the third approach, which can be termed the “Argentine model,” new workers have a pay-as-you-go tier combined with a private account tier: Costa Rica and Uruguay also adopted this model.

Until the early 2000s, Latin American countries were inclined to adopt some variant of these three approaches, with a tendency for later reformers to select the Chilean model, which gave new workers no choice but the personal accounts. Since then, pension reforms are as likely to eschew privatization entirely. Brazil, for example, has chosen to reform the parameters of its pay-as-you-go pensions, and Ecuador and Nicaragua have decided to postpone structural reforms.

Assessment of the fiscal, financial, and labor market effects of the reforms reveals that the results have been mixed. Performance in a variety of areas is assessed below.

Effect of Reform on System Balances

Fiscal imbalances were the primary motivation behind reforms, just as they appear to be the main concern in India, the Russia Federation, and the United States. The reforms seem to have had some success. Simulations by Gill, Packard, and Yermo (2005) for eight Latin American reformers indicate that the rate of
accumulation of pension debt fell sharply in most countries as a result of the reform (figure 1). In Bolivia, for example, pension-related debt would have been almost 160 percent of GDP in 2030 without reforms but is less than 50 percent of GDP with reforms. In Uruguay pension-related debt ratios for 2030 would be about 150 percent without reform and 70 percent with reform.

These are the long-term effects. In the immediate aftermath of reforms, however, these countries had to deal with the transition costs, as contributions were divided from paying benefits to the elderly to investing in the private accounts of workers. The promised benefits to current pensioners and older workers under the old system had to be paid, while part of the payroll tax flowing in had to be diverted to fund individual accounts. With contributions diverted into funded pension accounts, governments had to find ways of financing existing pay-as-you-go liabilities. For a variety of reasons, in some countries these transition costs proved to be higher than expected at the time of the reform. In Bolivia, for example, the pension-related deficit has been rising instead of falling, as projected.

More important is the fact that many countries had to finance the transition through increased government debt, some of it held by the new pension funds. With the investment regulations favoring government debt and with the thin capital markets in much of the region, Latin American workers essentially
swapped pay-as-you-go debt for government bonds. Fully two-thirds of the average investment portfolio consisted of government securities (figure 2). While a case can be made that this debt is more secure, the case of Argentina—where the government wrote down its debt by more than two-thirds—suggests that this greater security is a matter of degree. Moreover, with less than 20 percent of these funds going to corporate bonds and equities, the growth effects of the reform were also likely weak.

The reforms aimed at improving labor market efficiency by strengthening the links between contributions and benefits and by reducing the regressive transfers that characterized the previous social security systems. They appear to have been successful in reducing regressive transfers (figure 3).

Although reforms have addressed within-system equity concerns, the most inequitable aspect of unreformed social security systems in Latin America was that they excluded large shares of the population from even a semblance of income security. Reforms have been less effective in addressing this problem. While the closer links between benefits and contributions may have increased participation, the effect was small. Participation rates in most countries have essentially flat-lined at levels ranging from about 10 percent to about 67 percent of the active labor force (figure 4). This lackluster performance—despite closer links between contributions and benefits—can be attributed to high (and rising)
payroll taxes, as discussed below. But some of it also reflects the high management and insurance fees private pension providers charge.

**Effect of Reform on Financial Markets**

Reforms stimulated financial markets—but often at a high price for contributors. In Latin America as a whole the administrative costs of private schemes have been considerably higher than those of public schemes (Arenas de Mesa and Mesa-Lago 2006). In theory, private systems can reduce administrative costs through competition. In practice, multiple private providers lose the advantage of economies of scale, and considerable resources are spent on advertising and sales commissions. Administrators in private systems charge a commission (as a percentage of wages) for managing the old-age program plus a premium, transferred to an insurance company, to cover disability and survivor risks (Mesa-Lago 2006).

The evidence indicates that ensuring a captive clientele for these pension funds has led to the growth of a new industry in the reforming countries; assets held by pension funds more than doubled as a share of GDP between 1998 and 2004. But while financial market development has been hastened, contributors have generally paid a high price. This was especially the case in the early days of reform in countries such as Chile; for workers who had contributed for a decade or longer.
Figure 4. Pension System Participation Rates, before and after Reform, 1980–2000

Note: Reform dates are as follows: Argentina, 1994; Bolivia, 1997; Chile, 1981; Colombia, 1994; El Salvador, 1998; Mexico, 1997; Uruguay, 1996.
Source: Based on Gill, Packard, and Yermo (2005).

more than a quarter of contributions may have gone toward management and insurance fees. Even today, 15–20 percent of contributions go to management fees, and workers have to pay insurance fees as well (figure 5). Administrative costs may be even higher for low-wage workers in some countries. Low-wage workers often did not participate in the system before reform because of the weak connection between contributions and benefits; they may now find the new systems unattractive because of higher payroll tax rates and onerous administrative fees.

Administrative costs have generally come down as the pension fund industries have matured, but in some countries, such as Peru, the fees charged to contributors have not fallen commensurately. This has meant high profits for the fund managers. Between 1998 and 2002 the share of workers’ contributions going to fees remained steady in Peru, while fund expenses fell. As a result, profit rates skyrocketed (figure 6).

Chile’s recent experience appears to be similar. Competition among private pension providers in Chile has been limited because of the small number of administrators and high and increasing concentration among the largest funds (Arenas de Mesa and Mesa-Lago 2006). Even capable Latin American governments appear to find it difficult to effectively regulate these oligopolies.
Lessons for Other Countries

Not all of these findings are relevant for other countries, because the countries and the social security issues being debated are different. However, some general lessons do emerge.
**Effect of payroll taxes on labor market incentives.** While the move to private accounts had a small but positive effect on participation, the negative effect of higher payroll taxes may have been greater. Given near universal coverage of social security in developed countries and many transition economies, the discussion here is focused on the incentives to work rather than to contribute to social security. In particular, the concern is about the distorting effect of social security on the age of retirement.

The Latin American experience appears to support those who argue against raising the payroll tax rate. Two findings are noteworthy. The first is that there is little evidence from Latin America that, in the presence of high transaction costs, individual accounts led to stronger labor market incentives, as evidenced in participation rates. Payroll taxes went up in all countries except Chile and Uruguay (figure 7).

The second finding is that diverting payroll contributions from pay-as-you-go systems to individual accounts appears to have adverse fiscal implications that are far more potent. Most Latin American countries had to create space for the second pillar, which required that they downsize and redesign the first pillar. Chile, Bolivia, and Mexico provided a minimum pension guarantee to low-income workers whose personal accumulations fell below a specified amount. Many structural reformers in Latin America had to deal with financing the transition. One option was to increase taxation (through payroll or broader taxes, such as income tax or a general consumption tax) or borrowing (by issuing conventional

---

**Figure 7.** Payroll Tax Rates before and after Reform

![Payroll Tax Rates before and after Reform](chart)

*Source: Based on Gill, Packard, and Yermo (2005).*
public sector debt). Another was to reduce public spending—on pensions or in general or create new revenue (through privatization, for example).

When labor is mobile across sectors and the informal economy is large, structuring the premiums for social insurance programs as payroll taxes may be ineffective. Greater reliance on a broad tax base, such as an income or consumption tax instead of a payroll tax, may be more efficient. Using a broader tax is also more consistent with the poverty prevention and redistributive functions of the remaining public pooling pillar after introduction of the multipillar model, because such a tax reduces the wedge between the formal and informal parts of the labor force. The public pooling pillar enables individuals and households to manage shocks to their income should they need to, enabling them to be more enterprising.

**Effect of privatization.** The pension debate in Latin America has centered on two costs associated with the move to private accounts. One is the administrative fees charged by the special pension funds set up to manage these accounts and the costs of annuitizing the accumulated funds. The other is the fiscal costs associated with the transition to private accounts. Latin America’s experience with fiscal costs may be relevant for other countries.

The cases of Chile and Argentina provide contrasting experiences of the interaction of pension reform and fiscal effort. In Chile a strong fiscal effort characterized the lead-in to the pension reform: fiscal surpluses averaged more than 5 percent of GDP in the years before the 1981 reform, so that Chile’s fiscal deficits after the reform were mild and short-lived. In contrast, Argentina did not substantially bolster its fiscal situation in the years leading up to its 1994 reform. Though it ran small fiscal surpluses in the two years before the reform, there is reason to believe that its fiscal stance after the reform was worse than indicated by published figures. Payroll tax deductions reduced revenues and increased pension system deficits. About half of the deterioration of the consolidated public sector fiscal deficit between 1994 and the 2001 crisis was caused by the worsening social security balance.

The degree of protection against policy risk offered by privatizing a portion of mandated pensions may also be exaggerated. The experience in countries such as Argentina illustrates how any government-organized social security system—whether directly administered or simply mandated—can fall prey to politicians. Since the start of the system about half of all privately managed assets have been invested in government bonds. During the 2001 crisis, when the government forced the pension funds to swap dollar-denominated government bonds for peso debt, the share of government bonds in the private funds’ portfolio rose above three-quarters. Argentina is not unique in this regard. In Mexico, for example, several years after the reform the share of government bonds in pension fund portfolios is as high as in Argentina.
Pension system deficits contributed significantly to the deterioration of the fiscal balance in Argentina and other Latin American reformers. Many observers would underscore the importance of having a relatively strong fiscal position before undertaking structural reforms and of reducing the implicit debt of unfunded pay-as-you-go systems before making the debt explicit by shifting to a funded second pillar. But there are also concerns that this replacement could actually worsen the fiscal balances as reneging on explicit debt may be more costly than eroding real pension benefits (that is, reneging on implicit debt). The Latin American experience supports the views of advocates of general fiscal discipline rather than social security privatization as a prerequisite for ensuring a stable domestic financial sector and a friendly environment for private saving.

Worker objectives in buying old-age insurance. The experience in some Latin American countries raises questions about what individuals expect from their governments. Fiscal stability appears to be necessary for governments to fulfill these expectations; other conditions may also be required. Latin America’s experience can offer useful insights on how to curb rising pension costs and prevent pensioner poverty at the same time. Indeed, many developing economies already face rising pension spending, often combined with significant pensioner poverty (Barr 2006). Bourguignon et al. (2004) calculate the incidence of poverty among the elderly in 19 Latin American countries. Using household surveys to simulate the fiscal cost and impact on poverty rates of various uniform pension schemes, they show that a universal minimum pension would substantially reduce poverty among the elderly in all countries except Argentina, Brazil, Chile, and Uruguay, where minimum pension systems already exist.

Evidence from Chile and Peru, where the results of surveys (Gill, Packard, and Yermo 2005) designed to determine how households manage economic risk are available, reveals something about what workers expect from governments. At the time of the survey in Peru the government had not instituted a minimum pension guarantee; the survey revealed that private financial institutions were trusted more than all three branches of government. It also revealed that more risk-averse workers chose private funds over the reformed but still risky government pay-as-you-go option. In Chile the survey results were more revealing. Two decades after reform, workers seem to be using a system intended to act primarily as a vehicle for savings—with a small pooling component—mainly as a risk-pooling mechanism. Each cohort of workers that completes the minimum contribution requirements appears to be content with qualifying simply for the government’s guarantee of minimum pension—a modest means-tested amount of about 80–90 percent of the minimum wage.
Some researchers attribute this outcome to the moral hazard associated with low-income workers realizing that any contributions beyond this are a pure tax. In fact, this behavior occurs less among the working poor and more among middle- and higher income groups: it is more consistent with a desire to purchase some insurance against old-age poverty. The switch to other savings instruments after they have qualified for this insurance also indicates that workers see the mandated private accounts as relatively expensive or risky compared with other investments. There is evidence that other retirement investments—housing, household enterprise, even the education of children—in Chile are perceived as less risky than saving in the reformed pension system (Packard 2002). There is also evidence that households prefer to gain eligibility for the low, government-guaranteed annuity and continue to save outside the system, despite the variable but high real returns they could earn in the system. This evidence suggests that they may place greater value on security than on real rates of return (Gill, Packard, and Yermo 2005).

In countries where government is generally viewed as reliable, one could make the case that workers view the social security system more as a mechanism for insurance against poverty and less as a vehicle for saving to smooth consumption. The implication may be that the social security benefit structure should be made more progressive or the system made even more progressive than it is currently in developed countries.

**Main Policy Implications**

It is difficult to draw universal lessons on how to reform social security systems from the experience in Latin America. Developed countries already have universal coverage and well-developed financial markets, many developing economies outside Latin America do not have well-developed contributory social security systems, and transition economies in Europe and Central Asia face entirely different challenges than emerging markets in Latin America.

These differences notwithstanding, the Latin American experience provides some useful information about the behavior of (rational) workers, the responses of (profit-seeking) firms, and the responsibility of (fiscally constrained) governments. Put another way, the experience provides insights into how workers and firms react to changes in the structure of social security systems, what workers expect from their governments, and how governments can meet these expectations. The main policy pointers appear to be the following:

- *Keep payroll taxes low.* Strengthening the links between contributions and benefits can improve labor market incentives somewhat, but higher payroll tax rates will offset these benefits.

"Gill, Ozer, and Tutucu"
• *Keep benefits frugal.* Public pension benefits should be small and secure, in order not to unduly discourage saving for old age while providing insurance against poverty in old age.

• *Keep governments solvent.* Fiscal prudence is the most important rule for governments that wish to provide both a safe environment for private saving and reliable insurance against old-age poverty.

While these lessons emerge from the experience in Latin America, they are also consistent with fundamental principles of the economics of insurance. Ehrlich and Becker (2000) and others propose that optimal insurance implies that rarer and more idiosyncratic losses are better pooled, while frequent and more systemic losses should be saved for. This principle can be applied to the losses associated with old age. The blessing of rising longevity implies that losing the capacity to earn is an increasingly frequent loss for individuals. The blessing of falling poverty rates implies that being poor in old age is becoming an increasingly rare loss. Rising longevity necessitates a shift to self-insurance or saving as the way to smooth consumption over one's lifetime, while falling poverty implies a shift to market insurance or pooling.

The role of governments is to facilitate these actions by individuals to insure, self-insure, and self-protect. Since there are relatively few serious impediments to the ability of individuals to save for old age, the role for governments in encouraging saving for old age should be secondary and diminish. In contrast, in the case of poverty, because of the "social" nature of the loss being insured against and well-known problems with insurance markets, the role for governments is primary.

Much of the discussion in any country should center on the role of governments in helping individuals save and smooth consumption over their lifetimes and the need to help individuals insure against the losses associated with becoming destitute in old age. While it is clear that the mainstay for consumption smoothing should be individual saving, it is less clear what role the government should play in getting individuals to save. In the case of destitution in old age, however, the role for governments is clearer: it needs to provide an instrument for insurance against the increasingly rare loss associated with falling into poverty.

For various reasons social security systems have historically bundled these two functions. The implication is that as the role of government in saving is scaled down, the insurance function becomes more, not less, important. As Lindbeck and Persson (2003, p. 60) note in their cross-regional survey of social security reforms, "Reforms do not diminish the need for basic, or guaranteed, pensions. Quite the contrary; growing reliance on quasi-actuarial and actuarially fair systems which in themselves do not encompass any systematic intra-generational
redistributive elements, makes it even more imperative to maintain a safety net to prevent poverty in old age.”

Notes

Indermit S. Gill is the director of World Development Report 2009 at the World Bank; his email address is igill@worldbank.org. Ceren Ozer is an economist in the South Asia Poverty Reduction and Economic Management Unit at the World Bank; her email address is cozer@worldbank.org. Radu Tatucu is a junior professional in the East Asia Poverty Reduction and Economic Management Unit at the World Bank; his email address is rtatucu@worldbank.org. The authors would like to thank Jeffrey Brown, Peter Diamond, Andras Simonovits, and participants at a workshop at the 2006 American Economic Association meetings for useful comments, as well as three anonymous referees for many suggestions that improved this article.

1. Data were collected in specially designed surveys on risk, savings, and social insurance (Encuestas de Prevision de Riesgos Sociales) conducted in Santiago, Chile, in January 2000, and in Lima, Peru, in May 2002.

References


Why OECD Countries Should Reform Rules of Origin

Olivier Cadot • Jaime de Melo

With preferential trade agreements on the rise worldwide rules of origin—which are necessary to prevent trade deflection—are attracting increasing attention. At the same time, preference erosion for Generalized System of Preferences (GSP) recipients is increasing resistance to further multilateral negotiations. Drawing on different approaches, this article shows that the current system of rules of origin that is used by the European Union and the United States in preferential trade agreements (including the GSP) and that is similar to systems used by other Organisation for Economic Co-operation and Development countries should be drastically simplified if developed economies really want to help developing economies integrate into the world trading system. In addition to diverting resources for administrative tasks, current rules of origin carry significant compliance costs. More fundamentally, it is becoming increasingly clear that they are often designed to force developing economies to buy inefficient intermediate products from developed economies to “pay for” preferential access for the final product. The evidence also suggests that a significant share of the rents associated with market access (net of rules of origin compliance costs) is captured by developed economies. Finally, the restrictiveness of rules of origin is found to be beyond the levels that would be justified to prevent trade deflection, suggesting a capture by special interest groups. The article outlines some alternative paths to reforms. JEL codes: F13, F15

Rules of origin are an integral part of proliferating free trade agreements—countries belong to an average of six, according to a recent tally by the World Bank (2005, table 2.1)—and nonreciprocal preferential trade agreements such as the Generalized System of Preferences (GSP). Given the lack of progress on harmonization at the World Trade Organization (WTO) and given that regionalism is here to stay, rules of origin are likely to be increasingly important in the world trading system.
The primary justification for rules of origin in preferential trade agreements is to prevent "trade deflection," or taking advantage of low external tariffs or weak customs-monitoring capacities to bring in imports destined for more protected markets in a trading bloc (possibly after superficial conditioning or assembly). In effect, rules of origin are needed to prevent trade deflection for all preferential trade agreements short of customs unions, where trade deflection is not an issue because members have a common external tariff. Beyond the largely unimportant issue of tariff revenue, what is at stake is the unwanted extension of preferences to out-of-bloc producers, which would erode the value of those preferences to eligible producers. In preferential trade agreements between developed and developing economies, rules of origin are also sometimes justified on "developmental" grounds because they can help foster integrated manufacturing activities in developing economy partners.

However, this article provides evidence that, by their complexity, rules of origin impose substantial compliance costs on preferred producers. For instance, in addition to regime-wide rules, the European Union has more than 500 product-specific rules of origin (Cadot, de Melo, and Pondard 2006). As a result, these rules are increasingly difficult to observe. In the least developed economies the rules divert scarce customs resources from other tasks such as trade facilitation. In preferential trade agreements between developed and developing economies, forcing developing economy producers to source relatively inefficient intermediate goods locally or from developed economy partners rather than from the most price-competitive sources (as in, say, Asia) increases inefficiency and raises costs. The result is reduced value of preferences (compounding preference erosion in particular for least developed economies) and rent creation for developed country producers.

This potential for rules of origin to become a form of "export protection" was first observed by Krueger (1998) during negotiations for the North American Free Trade Agreement (NAFTA). It applies to all preferential trade agreements (including nonreciprocal preferential schemes) granted by Organisation for Economic Co-operation and Development (OECD) countries to developing economies. Moreover, there is an overwhelming evidence that this protectionist effect of rules of origin is not incidental but by design. Because rules of origin, unlike more traditional forms of trade protection such as voluntary export restraints or antidumping provisions, have so far largely escaped WTO disciplines: they are thus potentially a choice instrument for creeping protectionism.

New evidence reported in this article shows that the burden imposed by the rules of origin applied by the two main protagonists in preferential trade agreements, the European Union and the United States, is substantial whenever preferential margins are anything more than negligible. All told, the detailed evidence gathered here suggests that the current system of rules of origin applied by
developed economies is out of hand and defeats both the spirit of reforms aimed at bringing greater transparency to the multilateral trading system and the development-friendly intent of preference schemes.

In a recent communication, the European Union decided to consider simplifying its rules of origin. However, other OECD countries have so far refrained from reforming their rules and have opposed any discussion of reform of preferential rules of origin at the WTO. This article is a contribution to an overdue debate on how to design benign, transparent, and WTO-compatible rules of origin.

This article is organized as follows. The first section briefly recounts how product-specific rules of origin are defined in EU and U.S. preferential schemes and proposes an ordinal restrictiveness index summarizing their complexity. This index is shown to be correlated with EU and U.S. most favored nation tariffs (and thus, with the depth of trade preferences). The second section presents a simple framework for quantifying the costs associated with rules of origin: distortionary, administrative, and rent-transfer. The third section provides direct evidence of the effect of rules of origin on preference use and rent sharing using preference utilization rates and unit values. The fourth section qualifies the direct evidence by considering the Asian exception and the natural experiment provided by comparing the EU Everything But Arms initiative and the U.S. African Growth and Opportunity Act (AGOA), which have similar tariff-preference margins but different rules of origin. The fifth section provides further indirect evidence. The sixth section draws policy implications from the article’s findings and makes recommendations for simplifying existing rules of origin.

Rules of Origin: Definition and Measurement

Rules of origin in preferential trade agreements have two components: a small set of regime-wide rules and a large set of product-specific rules, typically defined at the Harmonized System six-digit level of disaggregation (HS-6). Both rules together are to ensure sufficient transformation. Because the European Union and the United States are the main users of preferential trade schemes among OECD countries, this article follows the approach of Cadot, de Melo, and Portugal-Perez (2005), describing briefly the rules for NAFTA, which have been in place for a long time and correspond closely to those applied by the United States in other preferential trade agreements, and those for the European Union’s “Pan-European system (PANEURO),” also called the “single-list” because it covers the common set of product-specific rules of origin that the European Union applies in all its preferential trade agreements (regime-wide rules differ across the European Union’s preference schemes such as the GSP or Cotonou Agreement). The analysis starts with regime-wide rules then turns to product-specific rules of origin.
Regime-wide Rules

Regime-wide rules usually include five components (these and other terms are defined in the glossary at the end of the article):

- A *de minimis* (or tolerance) criterion that stipulates the maximum percentage of nonoriginating materials that can be used without affecting the origin of the final product.
- A cumulation rule.
- A provision on whether "roll-up" applies.
- The status of duty drawbacks.
- The applicable certification method.

Table 1 describes how these regime-wide rules differ between the European Union and the United States.

<table>
<thead>
<tr>
<th>Preferential trade agreement</th>
<th>De minimis or tolerance rule</th>
<th>Absorption (roll up)</th>
<th>Cumulation</th>
<th>Drawback allowed</th>
<th>Certification method</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAFTA</td>
<td>7 percent (except agricultural and industrial products), 7 percent of weight for goods in chapters 50–63</td>
<td>Yes (except autos)</td>
<td>Bilateral</td>
<td>Not after 7 yrs</td>
<td>Self-certification</td>
</tr>
<tr>
<td>United States-Chile</td>
<td>10 percent (except agricultural and processed agricultural products)</td>
<td>Yes</td>
<td>Bilateral</td>
<td>Not mentioned</td>
<td>Self-certification</td>
</tr>
<tr>
<td>U. S.-GSP</td>
<td>10 percent, 10 percent of weight for goods in chapters 50–63</td>
<td>Not mentioned</td>
<td>Bilateral, limited diagonal</td>
<td>Not mentioned</td>
<td>Self-certification</td>
</tr>
<tr>
<td>Cotonou Agreement</td>
<td>15 percent</td>
<td>Yes</td>
<td>Full</td>
<td>Not mentioned</td>
<td>Two-step private and public and limited self-certification</td>
</tr>
<tr>
<td>EU GSP</td>
<td>10 percent (except goods in chapters 50–63)*</td>
<td>Yes</td>
<td>Bilateral, limited diagonal</td>
<td>Not mentioned</td>
<td>Two-step private and public and limited self-certification</td>
</tr>
</tbody>
</table>

*Note: Classification is carried out at the six-digit Harmonized System tariff line level. Each cell is the percentage of tariff lines that have the rules of origin in the corresponding row and in the corresponding column.

*Goods in chapters 50–63 (textiles and apparel) do not benefit from a *de minimis* provision.

Source: Cadot, de Melo, and Portugal-Perez 2005, table 1.
Even for regime-wide rules, table 1 gives the impression of “made-to-measure” rules. It also shows that regime-wide rules differ across preferential trade agreements for the same developed economy partner, confirming the hub-and-spoke characteristic of preferential trade agreements between developed and developing economies. Certification methods also differ between EU and U.S. preferential trade agreements; certification is easier to carry out in U.S. agreements, at least in principle, than in EU ones.

**Product-Specific Rules of Origin**

Devising methods for determining sufficient processing (or substantial transformation) has turned out to be very complex in all existing preferential trade agreements because the Harmonized System was not designed to define the origin of goods. Three criteria are used by the European Union and the United States to determine whether sufficient transformation has taken place in activities requiring processing (that is, anything but crude products):

- A change of tariff classification (at various levels of the Harmonized System), meaning that the final product and its imported components should not belong to the same tariff classification (in other words, that the local processing should be substantial enough to induce a change of tariff classification).
- A critical threshold for value added (in short, a value content rule).
- A specific manufacturing process (a so-called “technical requirement”).

For crude products the typical rule is “wholly obtained,” which permits no foreign content whatsoever, although other rules apply in special cases, such as fish products.

Both NAFTA (whose rules are also used in other U.S. preferential trade agreements) and PANEURO have a long list of criteria—including such technical requirements as the “triple transformation” requirement in textiles and apparel, which requires apparel to be woven from originating fabric and yarn. Criteria also include exceptions (making them more stringent) and allowances (making them less stringent). NAFTA relies more heavily on changes of tariff classification, though often in combination with other criteria. PANEURO relies mostly on value content and wholly obtained criteria, with wholly obtained criteria prevalent for GSP and African, Caribbean, and Pacific (ACP) exports of primary products with little processing.

As Krishna (2006) points out, when analyzing rules of origin, the devil is in the details because the complexity of rules of origin is what provides an opportunity for special interests to influence their design and administration. While many facets of rules of origin have been explored, rigorous empirical study of their
effects has been hampered by two difficulties, one relating to data on utilization rates, the other to measurement of the rules' restrictiveness.

First, data on preference utilization have been made freely available to the public only recently for the United States but not yet for the European Union (for example, Brenton and Manchin 2003 and the studies collected in Cadot, Estevadeordal et al. 2006).

Second, because rules of origin are a set of complex, heterogeneous legal rules, it has proved difficult to develop a reliable measure of their restrictiveness to serve as a synthetic indicator (much like effective rates of protection are a synthetic indicator of the restrictiveness of a country's trade regime). Estevadeordal (2000) has proposed an ordinal index of product-specific rules of origin restrictiveness (or R-index), taking values between one and seven, with higher values corresponding to more restrictive rules of origin. The index, constructed from a simple observation rule at the HS-6 level, where rules of origin are defined, is described below.

The observation rule is as follows (Cadot, de Melo, and Portugal-Perez 2005). Let CC stand for a change of chapter, CH for a change of heading, CS for a change of subheading, and CI for a change of item. A change of classification at the item level can be taken as less stringent than one at the subheading level, and so forth. So the criterion for classifying changes of tariff classification criteria is

\[ CC > CH > CS > CI \]  
(1)

But a change of tariff classification is often accompanied by one or two (in a few cases even three) additional requirements, such as value content rules, technical requirements, exceptions, or allowances. The observation rule assigns higher index values to changes of tariff classification when these requirements are added and lower ones in the case of allowances. For instance, a change of heading is given an index value of four, which rises to a five when accompanied by a technical requirement or exception but shrinks to three when accompanied by an allowance.

Though not amenable to quantification as effective rates of protection, the R-index plays the same analytical role; it is intended as an overall indicator of how trade-inhibiting the requirements that must be met by a product to obtain originating status. There is preliminary evidence that preferences have hidden compliance costs and that those compliance costs may be related to rules of origin. Table 2 shows evidence for the textile and apparel sector under NAFTA, the EU GSP, and the Cotonou Agreement (which grants tariff-free access for most ACP products to the EU market). Although NAFTA's and Cotonou's preference margins are equal, at 10.4 percentage points, their utilization rates vary widely: 50 percent for Cotonou compared with 79.9 for NAFTA. Cotonou's low rate of uptake despite deep preferences suggests hidden barriers. ACP countries benefit from full rather than diagonal cumulation (that is, intermediate purchases from
Table 2. Preferences and Utilization Rates for Textiles and Apparel

<table>
<thead>
<tr>
<th>Preferential trade agreement</th>
<th>Number of observations</th>
<th>Utilization rate (percent)</th>
<th>Preference margin (percentage points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAFTA (2001)</td>
<td>618</td>
<td>79.9</td>
<td>10.4</td>
</tr>
<tr>
<td>EU GSP (2004)</td>
<td>16,555 (HS-8)</td>
<td>52.2</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>12,920 (HS-6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cotonou Agreement (2004)</td>
<td>1,370 (HS-8)</td>
<td>50.0</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>1,175 (HS-6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Averages are unweighted. HS-6 is the six-digit Harmonized System level; HS-8 is the eight-digit Harmonized System level.

Source: Cadot, de Melo, and Portugal-Perez (forthcoming), table 3b.

Table 3. Preferences and Utilization Rates, All Goods

<table>
<thead>
<tr>
<th>Preferential trade agreement</th>
<th>Preference margin $\tau \geq 4%$</th>
<th>Preference margin $\tau \geq 8%$</th>
<th>Preference margin $\tau \geq 12%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>North American Free Trade Agreement$^b$</td>
<td>87 (1.239)</td>
<td>86.0 (558)</td>
<td>82.8 (287)</td>
</tr>
<tr>
<td>GSP$^c$</td>
<td>50.2 (1.297)</td>
<td>52.5 (91)</td>
<td>66.2 (44)</td>
</tr>
<tr>
<td>Cotonou Agreement$^c$</td>
<td>92.5 (1.627)</td>
<td>94.3 (892)</td>
<td>96.4 (566)</td>
</tr>
</tbody>
</table>

Note: Averages are unweighted. Numbers in parentheses are the number of tariff lines.

$^a \tau = (t^{\text{MFN}} - t^{\text{PREF}})/(1 + t^{\text{PREF}})$ is the preference margin.

$^b$Computed at the six-digit Harmonized System tariff-line level with 2001 data.

$^c$Computed at the eight-digit Harmonized System tariff-line level with 2004 data for 92 countries (GSP) and 37 countries (Cotonou Agreement) qualifying for preferential market access.

Source: Cadot, de Melo, and Portugal-Perez (forthcoming), table 2.

all partners qualify as originating) and a 15 percent tolerance rule compared with only 10 for the GSP, which also excludes the textile and apparel sector (chapters 50–63) from the 10 percent tolerance rule.

Table 3 shows that the evidence of hidden costs goes beyond the textile and apparel sector, where differences in uptake at similar margins may reflect composition effects. Define the preferential margin $\tau$ by the normalized difference between most favored nation and preferential tariffs

$$\tau_i = \frac{t_i^{\text{MFN}} - t_i^{\text{PREF}}}{1 + t_i^{\text{MFN}}}$$ (2)

Table 3 shows that, contrary to expectations, when the preferential margin rises, utilization rates fall for NAFTA. This suggests that an omitted variable is positively
Table 4. Tariff Peaks and the R-index

<table>
<thead>
<tr>
<th>Tariff peaks</th>
<th>Restrictiveness-index value</th>
</tr>
</thead>
<tbody>
<tr>
<td>North American Free Trade Agreement</td>
<td>( R = 6.2 ) (257)</td>
</tr>
<tr>
<td>Low tariffs</td>
<td>4.8 (1,432)</td>
</tr>
</tbody>
</table>

Total number of tariff lines 3,555 4,961

Note: Numbers in parentheses are numbers of tariff lines. Restrictiveness indexes are unweighted.

*a* Tariff lines whose tariffs exceed three times the GSP average.

*b* Tariffs lines whose tariffs are less than one-third of the GSP average.

Source: Cadot, de Melo, and Portugal-Perez 2005, table 3.

correlated with tariffs but negatively correlated with preference utilization. Rules of origin are an obvious culprit.

Table 4 shows that lines with tariff peaks (that is, with tariffs more than three times the average), where preferential margins are highest, do have higher \( R \)-index values than those with low tariffs. This relationship holds for both NAFTA and PANEURO.

Figure 1 confirms the patterns in tables 2–4; utilization rates do not really increase with tariff-preference margins. For NAFTA, they actually decrease due largely to the influence of the textile and apparel sector, where tariff preferences are deep and rules of origin stringent.

Quantifying the Effects of Rules of Origin

Although product-specific rules of origin, as already noted, take a variety of legal forms (changes of tariff classification, value content rules, technical requirement, and the like), they can all be represented conceptually as floors on domestic value added. Suppose that a producer in Madagascar wishes to sell a shirt under preferential access in the European Union, this shirt is made with both originating intermediate goods (that is, intermediate goods that are either local, EU-made, or imported from other qualifying countries, according to cumulation rules) and nonoriginating intermediate goods, say from Bangladesh, China, or India. Now assume that to satisfy origin requirements (whether change of tariff classification, value content rule, or technical requirement), the Malagasy producer uses a higher proportion of originating inputs than would be the case in the absence of rules of origin (which is precisely the rule's purpose).

Let superscript \( R \) denote a choice restricted by rules of origin. Unrestricted value added is \( v_{\text{un}} \), and restricted value added is \( v_{\text{R}} \), so rules of origin content
Figure 1. Average Utilization Rates for Different Preferential Margin Thresholds

<table>
<thead>
<tr>
<th>Average utilization rate (percent)</th>
<th>GSP</th>
<th>NAFTA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>43.7</td>
<td>86.1</td>
</tr>
<tr>
<td></td>
<td>50.2</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>56.8</td>
<td>86.4</td>
</tr>
<tr>
<td></td>
<td>52.5</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>71.4</td>
<td>84.8</td>
</tr>
<tr>
<td></td>
<td>66.2</td>
<td>82.8</td>
</tr>
<tr>
<td></td>
<td>46.8</td>
<td>81.9</td>
</tr>
<tr>
<td></td>
<td>45.4</td>
<td>77.8</td>
</tr>
<tr>
<td></td>
<td>55.9</td>
<td>78.5</td>
</tr>
<tr>
<td></td>
<td>54.7</td>
<td>76.6</td>
</tr>
</tbody>
</table>

Note: Cotonou Agreement includes 37 countries, computed at the eight-digit Harmonized System level; GSP includes 92 countries, computed at the eight-digit Harmonized System level, and the NAFTA includes 3 countries, computed at the six-digit Harmonized System level. Data are unweighted averages computed at the most disaggregated tariff-line level (table 2). Averages are based on more than 100 observations except for GSP (minimum of 27 observations for preference margins, \( \tau \), equal to or greater than 20 percent.).

Source: Cadot, de Melo, and Pondard 2006.

reduces to \( \nu_0^R > \nu_0 \), whether or not it explicitly takes the form of a value content rule. Thus, conceptually a value content rule can be thought of as a generic rule that can play the role of all others by quantifying the objective common to all. This principle is important because it underlies an approach to rules of origin reform, discussed later, that substitutes a value content rule—possibly, although not necessarily, at differentiated rates across products—for the current array of instruments. It also highlights how information on rules of origin restrictiveness can be aggregated across instruments and subsumed into a single restrictiveness index, which itself can be then aggregated across product lines by averaging.

Five results emerge from the quantitative analysis of the relationship between rules of origin restrictiveness and preference uptake:

For a given preference margin a higher restrictiveness index translates into a lower utilization rate, all other things being equal.

For a given restrictiveness index a higher tariff-preference margin translates into a higher utilization rate, all other things being equal.
The compliance decisions of individual firms are binary; how the decisions aggregate into industry-wide utilization rates depends on the unobserved distribution of compliance costs.

A lower pass-through of tariff preferences for the least developed economies (due to low bargaining power) implies lower uptake of preferences, all other things being equal.

Improvements in the uptake of preferences can be obtained either from reductions in the restrictiveness of rules of origin or from cost-reducing administrative simplifications (such as transparent and uniform criterion).

The third result implies that the statistical relationship between R-index values, preference margins, and utilization rates can only be “noisy” (that is, affected by a large unexplained component) at the aggregate (product-line) level. But notwithstanding the noise introduced by unobserved firm characteristics (which could be investigated only with firm-level data that are not currently available), figure 1 suggests an unambiguous relationship between preference margins, rules of origin restrictiveness, compliance costs, and utilization rates. It also suggests that, without a proxy for rules of origin restrictiveness such as the R-index, attempting to evaluate the effect of tariff-preference margins on the uptake of those preferences may lead to omitted-variable bias.

Keeping in mind that this framework captures only some of the effects associated with rules of origin, several observations are in order. First, administrative costs act as a technical barrier to trade; they result in resource waste, and in the welfare calculus of the effects of rules of origin they are more costly than the usual deadweight losses. Second, if costs are associated with certification, requests for preferential status would not be observed when preference margins are low. Third, compliance costs are particularly high for differentiated products, for which there can be quality as well as price differences between eligible (local) and noneligible intermediate goods. Because part of those costs is passed on to consumers in the countries that determine the rules of origin, high utilization rates does not necessarily imply that rules of origin have small effects.

Stiff rules of origin may inhibit or deflect trade altogether, not just the uptake of preferences. This was shown in the case of the Europe Agreements, free-trade agreements signed in 1991 between the European Union and the Central and Eastern European countries. Tumurchudur (2007a) showed that a large share of the exports from Central and Eastern Europe was deflected from EU markets by rules of origin, resulting in heavy losses. Evidence of trade-inhibiting effects is also apparent in the evolution of textile and apparel exports under AGOA and the Everything But Arms initiative, which is discussed in the exception and quasi-natural experiment section below.
Direct Evidence

In the absence of firm-level data Carrère and de Melo (2006) assume that the preference utilization rate for product line $i$ (the percent of exports sent under the preferential regime rather than the most favored nation one), referred to as $U_i$, rises with the tariff-preference margin, $\tau_i$ (which may be just equal to the most favored nation tariff when preferential access means tariff-free access) and shrinks with rules of origin compliance costs $c_i^R$. That is, $U_i = f(\tau_i - c_i^R)$ where $f(\cdot)$ is an increasing function, and, $c_i^R = (R\circ O_i)$, where $g(\cdot)$ is an increasing function (true compliance costs are firm-specific and are thus unobserved; all that is observed is the presence of $R\circ O_i$). These assumptions lead to an estimable relation of the form

$$U_i = \lambda + \alpha \tau_i + \sum_k \theta_k R\circ O_{ik} + \epsilon_i$$  \hspace{1cm} (3)

where $R\circ O_{ik}$ is a set of dummy variables indicating the presence of product-specific rules of origin (change of tariff classifications, exceptions, and so on). Results from estimating equation (3) on NAFTA data confirm that utilization rates rise with preferential margins and shrink in the presence of rules of origin (Cadot, de Melo, and Portugal-Perez 2005 for results using data for the European Union).

Carrère and de Melo (2006) combined their estimates with R-index values to compute an estimated ad valorem equivalent of total rules of origin compliance costs (administrative costs and costs due to higher input costs). Their estimates range from 3.5 percent for a change of chapter to more than 15 percent for combinations of rules of origin involving technical requirements. The strongly inhibiting effect of technical requirements appears to be an empirical regularity.

Even if the estimates are robust to a range of specifications, it is difficult to infer a sense of robustness from estimates derived from a relation like equation (3) because so much heterogeneity and so many "unobservables" influence preference uptake. Estimates have proved fairly sensitive to the inclusion of control variables, in particular when using EU GSP data.

An alternative is to restrict the analysis to products, for which the sole criterion used to determined origin is a value content. Drawing on the variation in EU value content criteria across product lines with value content the sole criterion, Cadot, Carrère, and Strauss-Kahn (2007) estimate an equation similar to equation (3), in which however the dummy variables for rules of origin are replaced with the continuous value content rule values. Using dummy variables for Harmonized System sections to control for heterogeneity across sectors and restricting the sample to tariff lines with substantial tariff-preference margins (above 2 or 5 percent), they find that utilization rates rise, all other things being equal, with the maximum foreign content allowed by the value content rule.
### Table 5. Estimated Effects on Preference Utilization and Rent Transfer of Relaxing a Value Content Requirement

<table>
<thead>
<tr>
<th>Preference margin</th>
<th>$\eta \leq 2%$</th>
<th>$\eta \geq 5%$</th>
<th>$\eta \leq 2%$</th>
<th>$\eta \geq 5%$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACP + GSP</td>
<td>ACP + GSP</td>
<td>GSP</td>
<td>GSP</td>
</tr>
<tr>
<td>Number of observations</td>
<td>19,261</td>
<td>5,958</td>
<td>13,448</td>
<td>4,305</td>
</tr>
<tr>
<td>Mean preferential margin ($\zeta$) (percent)</td>
<td>3.74</td>
<td>5.14</td>
<td>3.79</td>
<td>5.22</td>
</tr>
<tr>
<td>Mean utilization rate (percent)</td>
<td>0.12</td>
<td>0.17</td>
<td>0.17</td>
<td>0.22</td>
</tr>
<tr>
<td>Mean value content (percent of unit price)</td>
<td>58.8</td>
<td>58.2</td>
<td>58.9</td>
<td>58.6</td>
</tr>
<tr>
<td>Mean value of imports (euros)</td>
<td>1,475,182</td>
<td>2,376,301</td>
<td>986,979</td>
<td>159,694</td>
</tr>
<tr>
<td>Simulation: local content requirement reduced by 10 percentage points</td>
<td>Change in preference utilization rate (percentage points)</td>
<td>2.0</td>
<td>5.2</td>
<td>2.5</td>
</tr>
<tr>
<td>Total rent transfer from increased utilization (millions of euros)$^a$</td>
<td>21.7</td>
<td>37.4</td>
<td>25.3</td>
<td>11.6</td>
</tr>
</tbody>
</table>

$^a$Evaluated at the mean value of imports.

Source: Authors' computations based on Cadot, de Melo, and Portugal-Perez (forthcoming), table 6.

Since a single value content criterion is a serious candidate for reform, at least in the case of the European Union (Stevens et al. 2006 and Cadot, de Melo, and Pondard 2006), table 5 reports two illustrative simulations based on these estimates. The mean local content requirement is 58 percent and preference margin 3–5 percent depending on the sample; mean utilization rates are rather low—between 12 and 22 percent. The bottom of the table shows the first-round effects (no supply response) of reducing the local-content requirement by 10 percentage points. Utilization rates rise by 2–5 percentage points (row 6), raising the rent transfer by €21–37 million, for a mean value of imports of €1.5–3.0 billion.

To fully grasp the welfare effects of rules of origin, the rent distribution between the exporting and importing country must be factored in. This implies estimating the pass-through effect of tariffs on consumer prices (that is, the extent to which preferences translate into a higher producer price for exporters). Estimates for AGOA preferences (Olareaga and Özden 2005) and for the Caribbean Community ( Özden and Sharma 2006) are that between one-third and one-half of tariff reductions are passed on to producers.

However, part of the border-price increase could reflect the compliance costs discussed above. Using a monopolistic-competition model with differentiated products in which Mexican exporters can export product $j$ either to the rest of the world (under most favored nation status, at price $p_{y}^{M}$) or to the
Table 6. Exports, Unit Costs, and Prices under Preferential Market Access and a Binding Minimum Local Content Requirement

<table>
<thead>
<tr>
<th>Simulations</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)*</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference margin (percent)</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Administrative unit costs (percent of unit price)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Unconstrained, minimum local content requirement (percent of unit price)</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>Constrained, minimum local content requirement (percent of unit price)</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>Preferential exports (percent change from scenario with no preferential access)</td>
<td>15.9</td>
<td>11.1</td>
<td>1.15</td>
<td>7.1</td>
<td>10.7</td>
</tr>
<tr>
<td>Unit costs (percent change from scenario with no preferential access)</td>
<td>0</td>
<td>1.9</td>
<td>6.7</td>
<td>1.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Unit net price (percent change from scenario with no preferential access)</td>
<td>2.9</td>
<td>3.0</td>
<td>3.4</td>
<td>2.2</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Note: Unit net price set equal to 1, initial output to 100, and value-added to 20. All output is exported 40 percent to preference-receiving destination. For columns 1–4 nonoriginating inputs are set to 75 percent of intermediate good input purchases. This implies that initial (unconstrained) local content is 20 + 0.25 × (80) = 40. Setting the minimum local content requirement at 50 percent implies reducing nonoriginating intermediate goods to 62.5 percent of intermediate good purchases. For column 5 nonoriginating inputs are set at 80 percent and reduced to 75 percent through the minimum local content rule.

*Same as column 2 but with low value for the elasticity of substitution between originating and nonoriginating materials (0.5 instead of 2).

Source: Authors’ computations adapted from model in Cadot et al. 2005.

United States (under NAFTA, at price $p_N^m$), Cadot et al. (2005) estimate the following relationship

$$\text{NAFTA markup} = \alpha_0 + \alpha_1 \tau_j + \alpha_1 CC_j + \alpha_3 TECH_j + \varepsilon_j$$

where “NAFTA markup” is the percentage by which Mexico’s NAFTA shipment prices are raised over comparable most favored nation shipment prices, CC$_j$ is a dummy variable marking a change of tariff classification at the chapter level, and TECH$_j$ is a dummy variable marking a change of a technical requirement.

When estimated at the HS-8 level, equation (4) is the best tool to compare prices in different markets. With complete pass-through ($\mu = 1$ in equation (4)) the estimated coefficient for $\alpha_1$ would be close to one, but Cadot et al. (2005) find it substantially below one. They also obtain negative and significant estimates for ($\alpha_2$, $\alpha_3$) indicating that rules of origin costs are at least to some extent passed on to consumers. Once rules of origin are taken into account, the backward
pass-through of preferences to producer prices falls from 80 percent of the margins to only 50 percent. They also show, using input–output links, that U.S. producers of intermediate goods are able to retain a substantial share of the rents generated by rules of origin downstream. That is, stiff rules of origin on, say, Mexican shirts exported to the United States significantly raise the price of fabric exported by the United States to Mexico for use in those shirts. This reflects the fact that rules of origin create a captive market for U.S. intermediate goods.

An Exception and a Quasi-Natural Experiment

The covariation of utilization rates and margins does not account for all the effects of rules of origin. Case studies such as those reported in Cadot, de Melo, and Pondard (2006) and Stevens et al. (2006) provide useful complementary evidence, although they conclude that each case is different, thereby explaining if not justifying the current maze. An exception and a quasi-natural experiment are drawn here, with both suggesting that rules of origin are, as they stand, unnecessarily restrictive.

Asian Exception

In a world where rules of origin are as cumbersome and complicated as they are (Estevadeordal and Suominen 2006 for a detailed description), Association of Southeast Asian Nations (ASEAN), Free Trade Area (AFTA) and the ASEAN-China Free Trade Area (ACFTA) stand out as exceptions. To obtain originating status (that is, to fulfill the criterion of sufficient processing), either the wholly obtained criterion (for a few agricultural products) or a single-value content rule requiring 40 percent local content (for most products) is used. This rule has been relaxed by allowing a choice between criteria for countries that found it too constraining. For instance, under ACFTA the importer can choose a change of tariff classification can be used as an alternative to the 40 percent local content for obtaining origin for leather goods, and some specific process criteria are also accepted for some textile products.7

So why are rules of origin under AFTA less stringent than elsewhere? First, until recently Asian regionalism was more about cooperation than about preferential trade. Under the aegis of the United States, Asia-Pacific Economic Cooperation was set up specifically to avoid preferential trade and the formation of an Asian trade bloc. Much of the region's integration in the world economy has been driven by unilateral tariff reductions. Second, regional trade has made possible the rise of the Asian manufacturing matrix in which labor-intensive stages of production initially carried out in Japan—and later in the Republic of
Korea—were outsourced to the region’s lower wage countries. The resulting regional production networks have contributed to the price-competitiveness of Asia’s exports, which has benefited the whole region. Stiff rules of origin would have jeopardized this successful model.

This “Asian exception” has been conducive to the successful development of Asian countries that have fully participated in “verticalizing” trade (the development of cross-border supply chains generating trade in intermediate products). In this unusual setup (relative to other global trading patterns), intraregional trade in politically sensitive final products where protection is highest was insignificant. Thus, the political-economy forces that would usually lead to the complex rules of origin observed elsewhere have not been at work so far. As a result, low-income countries such as Cambodia and Lao PDR have been able to participate in the fragmentation of production according to comparative advantage.\(^8\) Arguably, Asia’s simple and uniform rules of origin requirement is an example of the kind of rules of origin that would really be development-friendly.

**AGOA and Everything but Arms: a Natural Experiment**

In the textile and apparel sector, the choice area for obscure and trade-inhibiting rules of origin, the one notable exception is the U.S. preferences granted to 22 Sub-Saharan African least developed economies under AGOA. Thus, comparing African apparel exports to the European Union and the United States provides a quasi-experimental situation in which the effects of rules of origin on the uptake of trade preferences are analyzed. This quasi-experimental situation, first studied by Brenton and Özden (2005), comes from the combination of different rules of origin with very similar rates of preference margins (textiles and apparel receive approximately the same protection in the EU and U.S. markets. In 2001 the EU-15’s most favored nation tariff was 10.1 percent compared with 11.7 percent for the United States, and duty-free access applied to both Everything But Arms eligible and the 34 AGOA-eligible African countries).

To qualify for preferential access to the U.S. market, an exporter must prove that the garments are produced, cut, and sewn in the area benefiting from preferential access (here, AGOA). Cotton products must be made from originating fabric, yarn, and thread, with diagonal cumulation somewhat relaxing the requirement, since fabric originating in other member countries qualifies. However, this rule, known as “the triple transformation” rule, was relaxed for 22 least developed economies under AGOA’s “special regime,” which permits the use of third-country fabric.\(^9\) That is, the special regime reduces the transformation requirement to a single transformation (from fabric to garment).

Fifteen of AGOA’s special regime beneficiaries are also eligible for the European Union’s Everything But Arms Initiative. But no such relaxation applies to exports...
to the European Union under either the Cotonou Agreement or Every But Arms preferences. EU rules of origin for apparel require production from originating yarn, which implies a "double transformation" from yarn to fabric and from fabric to clothing. The European Union's "double-transformation" rule obviously makes compliance difficult for countries that have no textile industry. Small or poor countries that cannot profitably produce fabric— weaving is a capital-intensive activity involving expensive machinery, particularly for woven products—should not, from an economic-efficiency viewpoint, set up the vertically integrated local value chains that would satisfy the double-transformation rule.

In apparel preference utilization rates are very high under both AGOA (97.36 percent in 2004) and Everything But Arms/Cotonou (94.9 percent). Cotonou has rules similar to those that Everything But Arms has for apparel. However, export volumes evolved quite differently for the 15 least developed economies that benefit from both schemes. Figure 2 shows a substantial increase in the value of apparel exports with AGOA's entry into force in 2000 (in particular for Lesotho and Madagascar). By contrast, the value of exports from this same group of countries did not rise following the adoption of Everything But Arms—in fact it fell slightly. Of course, the exports that remained flat for those countries should come as no surprise since they already benefited from Cotonou preferences, which give almost

---

**Figure 2.** Apparel Exports of 22 Countries Benefiting from the AGOA Special Regime, 2004

![Graph showing apparel exports](image_url)

**Note:** Benin, Botswana, Cameroon, Cape Verde, Ethiopia, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Swaziland, Tanzania, Uganda, and Zambia. *Botswana, Cameroon, Ghana, Kenya, Lesotho, Madagascar, Namibia, Nigeria, and Swaziland.*

**Source:** Portugal-Perez (2007) based on the WTO Integrated Data Base

---

*The World Bank Research Observer, vol. 23, no. 1 (Spring 2008)*
as much access as Everything But Arms (with slightly more lenient rules on cumulation). In effect, nothing changed for them on this front, and along with other ACP countries they largely continued to request access under Cotonou, with which they were familiar, rather than Everything But Arms. But AGOA’s special regime did not merely trigger a catch up of U.S.-bound exports toward already high levels of EU-bound exports: it dwarfed them. Thus, unlike AGOA’s special regime neither Cotonou nor Everything But Arms appeared to have offered a preference mix (tariff preferences and rules of origin) conducive to export growth.

Because the data in figure 2 are computed at the HS-6 product level, it is safe to assume that heterogeneity in export composition is largely controlled for. This is confirmed by formal econometric evidence. In a model that controls for differences in preference margins and for demand shifters in the EU and U.S. markets, Portugal-Perez (2007) finds that relaxing rules of origin for apparel (captured by a dummy variable corresponding to the introduction of the AGOA’s special regime) raised apparel exports significantly for beneficiary countries. Because the special regime was not introduced in the same year for all countries, its effects are well identified statistically, and Portugal-Perez’ results strongly suggest that the difference in performance apparent in figure 2 is indeed attributable to differences in rules of origin regimes.

AGOA’s special regime seems to have encouraged growth not only at the “intensive margin” (higher volumes) but also at the “extensive margin” (diversification by addition of new products). As new products were exported to both countries (an active extensive margin), the rate of increase in new products was several orders of magnitude higher for the U.S.-bound goods than for EU-bound ones, which is an important achievement. Product diversification is one measure of industrialization, particularly at early stages of the economic development process (Cadot, Carrère, and Strauss-Kahn 2007 and references therein). Controlling for other factors, countries that have a more diversified industrial base enjoy less volatile growth and are better poised to absorb shocks. Only three countries in Sub-Saharan Africa—Lesotho, Madagascar, and Senegal—export more than 50 products to either the European Union or the United States. Thus, if the development objective of rules of origin is to be taken seriously, encouraging export growth at the extensive margin is important, and in this regard Everything But Arms and Cotonou’s performance are again disappointing compared with that of AGOA’s special regime.

Taken together, the brief discussion here on the Asian exception and the comparison of AGOA with the Everything But Arms initiative suggests two results:

Limited differences between preferential regimes can have drastic effects on their performance: AGOA’s relaxation of the triple transformation rule gave a significant boost to Sub-Saharan African apparel exports.
Utilization rates are an incomplete measure of the performance of preferential regimes, as the inhibiting effect of stiff rules of origin can be felt on trade volumes as well.

Indirect Evidence

Taking inspiration from the early work by Herin (1986) for EFTA, Cadot, de Melo, and Portugal-Perez (forthcoming) applied revealed-preference arguments to estimate upper and lower bounds of compliance costs. Arguably, this nonparametric approach could be more robust than the parametric evidence reported above. By revealed preference, for products with 100 percent utilization rates the net benefit of preferences is positive for all firms. Since everyone uses the preferences, the ad valorem equivalent of compliance costs cannot be larger than the tariff-preference margin. Conversely, for products with zero percent utilization rates, since no one uses the preferences, the compliance cost cannot be smaller than the preference margin.

For remaining sectors (those with utilization rates between 0 and 100 percent) the story is more complicated because of firm heterogeneity, so assumptions must be made. Cadot, de Melo, and Portugal-Perez (forthcoming) argue that, firm heterogeneity notwithstanding, the average exporter (in terms of compliance costs) is not too far from indifference between the preferential and the most favored nation regimes, which means that the compliance cost is about equal to the tariff-preference margin. Applying this reasoning gives trade-weighted ad valorem estimates of 4.7–8.2 percent depending on sectors for PANEURO and 1.8–1.9 percent for NAFTA—values in line with the econometric estimates of Carrère and de Melo (2006) reported earlier.

How then should requests for preferential status be interpreted when tariff preferences are nil? Beyond (likely) errors in data transcription, the logical possibility would be that administrative costs are negligible, but this contradicts the evidence (the nonparametric approach described in the previous paragraph gave estimates of pure administrative costs slightly above 3 percent in ad valorem form). Francois, Hoekman, and Manchin (2006) elegantly addressed this problem by modeling the determinants of utilization rates for EU trade with ACP countries in a switching-regression framework where the relationship between the variable of interest (utilization rates) and explanatory variables varies between two regimes: one for low-margin sectors and the other for high-margin ones. The dividing point between the two regimes is determined by the data using an algorithm developed by Hansen (2000). They found that exporters start requesting preferences when preferential margins are in the 4.0–4.5 percent range, a result that
is also broadly consistent with the nonparametric estimates of compliance costs reported above.

Other studies using aggregate bilateral trade data also suggest costs associated with the presence of rules of origin. Using a gravity model of bilateral trade, Anson et al. (2005) find that after controlling for the other determinants of the volume of bilateral trade, including the presence of free trade agreements, the intensity of bilateral trade is inversely related to the values taken by the $R$-index. Using a similar framework, Augier, Gasiorek, and Tang (2005) find that the volume of bilateral trade is lower when cumulation is on a bilateral rather than a full basis, leading them to suggest that rules of origin should be relaxed to allow for full cumulation.

The evidence reported so far in this article is overwhelming: rules of origin are burdensome and foster economic inefficiency. But this article also argues that they have a role in combating trade deflection, so calling them trade barriers is not enough. To make progress in designing “clean” rules of origin, a key part of the argument is to tell apart, in their current characteristics (and in particular their restrictiveness), how much is attributable to their antideflection role compared with how much is simply capture by special interests. Portugal-Perez (2006) tries to address this issue by decomposing variations in the $R$-index into a component attributable to trade deflection and one associated with lobbying or political-economy motives. He estimates this decomposition for Mexican textile and apparel exports to the United States under NAFTA using the following equation

$$RoO_i = \beta[TD_i] + \gamma[PE_i] + \varepsilon_i$$

(5)

where $(RoO_i)$ is $R$-index values at the HS-6 level. The regressors are the trade deflection vector, which includes a proxy for the extent of product differentiation (the more homogeneous the product, the more there is to gain from arbitraging even small differences in external tariffs), and differences in external tariffs (the larger these differences the more there is to arbitrage). Political-economy variables including the level of the United States' most favored nation tariff (a proxy for lobbying power) revealed comparative-advantage indexes and the value of Mexican exports to the rest of the world (a proxy for potential penetration of the U.S. market).

Portugal-Perez finds strong and quite robust correlations, suggesting that both sets of factors are at work in explaining cross-sectoral variations in rules of origin restrictiveness. Using estimated parameter values, he constructs a counterfactual distribution of $R$-index values across goods in the absence of political-economy correlates (that is, by setting $\gamma = 0$ in equation (5). The two distributions (actual and counterfactual) are reported in figure 3. They show that political-economy concerns (which shift the actual distribution to the right of the counterfactual) contribute to the overall restrictiveness of rules of origin. Drawing on the
estimates discussed earlier by Carrère and de Melo (2006), he concludes that capture by special interests may have raised the costs of rules of origin an average of 3.5–11 percent of good value, a very steep increase in the face of the shallow preferences that are generally granted.

Simulation methods provide another way of obtaining orders of magnitude of rules of origin effects on trade. Francois, Hoekman, and Manchin (2006) use their estimate of compliance costs to simulate the effects of trade liberalization by developed economies on low-income countries in a multiregional trade model. Despite preference erosion, low-income countries gain instead of losing from trade liberalization by the European Union because the “rectangle” deadweight losses associated with compliance costs are eliminated.

Table 6 provides alternative estimates from a partial-equilibrium perspective, taking as an example a GSP country benefiting from a 10 percent preferential margin in the EU (or U.S.) market (row 1) but forced to raise its minimum local content from the value in row 3 (40 percent, except in column 5) to the value in row 4 (50 percent, except in column 5). When present, administrative costs, also expressed as a percentage of the unit price, are given in row 2. The table’s bottom three rows show the effect of rules of origin on equilibrium exports and prices.

Column 1 shows the benefits that accrue to the GSP producer from receiving a 10 percent preference margin with no constraint on the sourcing of inputs. For this constellation of elasticity (all are on the high side to reflect the likelihood
that products from different origin are close substitutes, whether at the intermediate- or final-good level), the pass-through is 2.9 percent (row 7) out of a preference margin of 10 percent, in line with econometric estimates mentioned in the section on direct evidence. Exports increase by 16 percent, but costs do not increase because inputs are bought at constant world prices.

Column 2 shows what happens when the producer must reduce the use of nonoriginating materials to meet a value content rule of 50 percent (a 25 percent increase from column 1). For the example, where value added is 20 percent and unconstrained purchases of nonoriginating intermediate goods equal 75 percent of the value of total intermediate good purchases, raising the minimum local content from 40 to 50 percent implies that purchases of nonoriginating intermediate goods must be reduced to 62.5 percent. The result of forcing producers to shift away from preferred intermediate goods is a higher unit production cost resulting in lower export volume, with the 1.9 percent increase in unit cost passed on to EU and U.S. consumers. Matters get worse if substitution possibilities for materials from different origins are low (column 3), which might be the representative of industries with a lot of transformation and many production stages. \(^{11}\)

Column 4 mirrors column 2 but adds administrative compliance costs of 2.5 percent. This further penalizes the GSP producer, even though part of this cost increase can again be passed on to consumers in the importing country. Of course, if GSP producers were competing with close substitutes, they would be unable to pass on the price increase. Finally, column 5 considers a simulation that might be fairly representative of an industry with enough originating intermediate good purchases that the shift to a 40 percent minimum local content would not affect producers much. In this case, the net price to producers might go up by about one-third of the preference margin, resulting in a modest supply response of about 10 percent.

**Implications for Reform**

If rules of origin are a legitimate way to prevent trade deflection by mandating that sufficient processing take place in the preferential zone, the accumulated evidence reported in this article indicates that they have gone vastly beyond that role, becoming akin to technical barriers to trade. Various estimates suggest that the compliance costs associated with meeting origin requirements in preferential trade agreements range between 3 and 5 percent of final product prices—a very stiff price tag for preference margins that are often thin, given that most favored nation tariffs are low in most sectors except textiles and apparel. Controlling for preferential margins, utilization rates are lower in product lines with more
restrictive rules of origin and when producers are limited in the sourcing of their intermediate good purchases.

Because of their trade-inhibiting effects, rules of origin hinder the integration of preference-receiving least developed economies in the world economy and thus work at cross-purposes with the development-policy goals of EU and U.S. preferences. For Sub-Saharan African countries supplying apparel products to the European Union, even high utilization rates hide obstacles to export growth caused by the double-transformation requirement imposed on those products.

This article also shows that in the case of the European Union and the United States, the two largest users of preferential trade agreements, rules of origin are stricter for products with tariff peaks where preferences could be most valuable. The correlation between the presence of tariff peaks and that of highly restrictive rules of origin suggests capture by protectionist interests, a hypothesis largely confirmed by political-economy theory and evidence. Moreover, because rules of origin have so far escaped WTO disciplines—whereas other, more traditional trade-policy instruments are brought under increasingly stringent ones—they stand as a choice candidate for creeping protectionism.

Despite the prevalence of capture by special interests, two quasi-natural experiments point to broad directions for reform. First, the relaxation of the U.S. triple-transformation requirement in textile and apparel for Sub-Saharan African producers under AGOA has proved to strongly encourage export diversification and growth compared with exports destined to the European Union, which are subject to stricter rules under the Everything But Arms initiative (which otherwise features similar preference margins). Second, low-income Asian countries operating under simple and benign rules of origin have been able to rapidly integrate themselves into cross-border supply chains and have, as a result, tremendously benefited from the verticalization of world trade.

These observations suggest that a multilateral agenda for preferential rules of origin reform, a key step in bringing preferential trade agreements under WTO disciplines, would have to move along three dimensions: harmonization, simplification, and relaxation. Harmonization between trading blocs, although unlikely to be attained anytime soon, is desirable in view of the “spaghetti bowl” of preferential trade agreements and is a prerequisite for simple and mutually consistent cumulation rules. The European Union has set an example in this regard with the PANEURO system, designed precisely to facilitate cumulation across preferential zones.

For simplification arguments in favor of a single across-the-board rule are much like those in favor of uniform tariffs—that is, simplification fosters transparency and mitigates capture. Clearly, technical requirements should be targeted for elimination first because they are the most opaque, difficult to harmonize, and capture-prone instruments. Leaving aside agricultural products that could still
operate under the wholly obtained criterion and keeping in mind that any uniform rule will affect industries and countries differently, two avenues could be considered: a simple change of tariff classification, say at the subheading (HS-6) level so that it is not too restrictive or a uniform value-content rule.

Some information can be gleaned in this regard from the European Union’s recent review. The change of tariff classification has the advantage of simplicity, transparency, and low administrative costs. But the Harmonized System tariff nomenclature was designed to collect trade statistics, not to separate products and confer origin, so defining the change of tariff classification at a uniform level would produce erratic results across sectors. This would call for exceptions to uniformity, opening up a Pandora’s Box of special deals. Moreover, a change of tariff classification that would not easily lend itself to differential treatment for least developed economies should be an objective (see below).

Notwithstanding conceptual clarity, a value content rule may be less than straightforward to apply in practice.\(^{12}\) It may increase producer risk due to the sensitivity of costs to exchange-rate, wage, and commodity-price fluctuations and is also burdensome to apply for customs officials. However, it is simple to specify and transparent, and it allows for differential treatment of least developed economies. All told, if properly specified, it is the best candidate for an across-the-board criterion, ideally in combination, at the exporter’s choice, with a change of tariff classification. In this spirit Tumurchudur (2007b) estimated for each good the maximum foreign content that would make a value content rule equivalent to the current array of NAFTA’s rules of origin. Her method consisted of three steps.

First, she estimated the statistical relationship between utilization rates and rules of origin, including value content rules. Second, she inverted that relationship to find the rate of a value content rule that would give a utilization rate equal to the current one. Third, she calculated the trade-weighted average of that maximum content. This neutral average turns out to be a very low 21 percent of the good’s value in maximum foreign content, confirming the diagnosis that NAFTA’s rules of origin are very restrictive. More important, this rate provides a transparent and fully comparable benchmark which is to base discussions of reform and harmonization.

If the slow pace of harmonization talks at the WTO is any indication, the reform agenda described above may be overambitious by several orders of magnitude; even if the European Commission manages to complete the agenda, competition between systems may trigger similar rounds of simplification elsewhere, including in free trade agreements between developing economies in Africa and Latin America, whose rules of origin are often directly inspired by NAFTA and PANEURO. However, the outcome of the EU reform process is highly uncertain at this stage; moreover, even if the plan to adopt an across-the-board value content
criterion survives, it is not clear that the rate of this value content rule would be uniform. Nor is it certain (perhaps even less) that it would relax the restrictiveness of the current system.

More immediate, win–win steps may be a better way to proceed. A simple first step would consist of eliminating rules of origin requirements for tariff lines with preferential margins below 3 or perhaps even 5 percent (the rate could be agreed upon in the context of multilateral negotiations at the WTO). This would be an all-around winning proposition since resources would be freed for other purposes, especially in developing economies, but also for consumers in developed economies, who would no longer bear part of the increased costs associated with compliance. A second step would be to allow for differential treatment not across sectors, but across beneficiaries, with low value content requirements for least developed economies reflecting the empirical observation that the "slices" of value added in least developed economies reflecting the empirical observation that the "slices" of value added in least developed economies through cross-border production networks are generally thin. In this regard, the experience with the U.S. special regime granted in textile and apparel to African producers under AGOA is most encouraging.

Appendix. Glossary of terms

**Harmonized System.** A system of classification for traded goods in which all countries belonging to the World Customs Organization participate. It classifies traded goods into (by increasing order of disaggregation) 21 sections (one digit), 99 chapters (two digits), 1,417 items (four digits), and 4,998 subitems (six digits). Beyond that (eight- and ten-digit), classification systems are no longer harmonized across countries and are subject to frequent classification changes.

**Preference Margin.** The difference between most favored nation and preferential tariffs.

**Preference Pass-Through.** The percentage of a tariff-preference margin that is "appropriated" by exporters in the form of an increase in the export price. It is inversely related to the bargaining power of importers.

**Preferential Status.** Whether a good is eligible for the preferential tariff rate.

**Technical Requirement.** Rule of origin that imposes a certain type of production process or the use of certain specified technology or standard.
Trade Deflection. Use of the country with the lowest external tariff by importers in a free trade agreement (which reduces tariff revenue for others). This notion is distinct from Vinerian "trade diversion."

Utilization Rate. Share of exports shipped under the preferential (as opposed to most favored nation) regime.

Regime-wide Rules of Origin

Absorption or Roll-up. Principle that allows nonoriginating materials that have acquired origin by meeting specific processing requirements to maintain this origin when used as input in a subsequent transformation. In other words, the nonoriginating materials are no longer taken into account in calculating value added. The roll-up or absorption principle is used in most preferential trade agreements (in particular, the EU GSP and the Cotonou Agreement), although a few have exceptions for the automotive sector.

Cumulation. Principle that allows producers from one member country in a preferential trade agreement to import nonoriginating materials from another member country without affecting the final product’s originating status. There are three types of cumulation rules: bilateral, diagonal, and full Bilateral cumulation. It is the most common type and applies to trade between two partners in a preferential trade agreement. It stipulates that producers in country A can use inputs from country B without affecting the final good’s originating status as long as the inputs satisfy the area’s rules of origin.

Diagonal Cumulation. Under diagonal cumulation (the basic principle of the EU’s PANEURO system), countries in a preferential trade agreement can use materials that originate in any member country as if the materials originated in the country where the processing is undertaken. Under full cumulation all stages of processing or transformation of a product within countries in a preferential trade agreement can be counted as qualifying content regardless of whether the processing is sufficient to confer originating status to the materials themselves. Full cumulation allows for greater fragmentation of the production process than bilateral and diagonal cumulation.

Duty Drawbacks. Refunds to exporters of tariffs paid on imported intermediate good inputs. Many preferential trade agreements, especially in the Americas, mandate the elimination of duty-drawback schemes for exports to partner countries on the grounds that a duty drawback claimed by a producer in country A to export to country B would put that producer at a competitive advantage compared with
domestic producers in country B given that the producer in country A already benefits from the elimination of intrabloc tariffs. Eliminating duty drawbacks as part of a preferential trade agreement can harm the profitability of final-good assembly for export to partner countries in the area, although tariff escalation, when present, already provides some protection for final-assembly operations (because it implies lower tariffs on intermediate goods than on final ones).

**Product-Specific Rules of Origin**

*Allowance.* An amendment to a mandated change of tariff classification that excludes some categories from noneligibility (that is, a final good belonging to, say, chapter 11 can embody imported inputs belonging to any other chapter or from chapter 11 itself but between headings X and Y).

*Change of Tariff Classification.* Rule of origin requiring that a final good made with imported inputs belong to a Harmonized System category that differs from that of its imported inputs (as proof of transformation). The mandated change of tariff classification can be specified at the chapter (two digits), heading (four digits), subheading (six digits), or item (eight digits) level.

*Exception.* An amendment to a mandated change of tariff classification that excludes some categories from eligibility (that is, a final good belonging to, say, chapter 11 can embody imported inputs belonging to any other chapter except headings X to Y).

*Value Content.* Rule of origin requiring a minimum percentage of local value (materials or value added) or a maximum percentage of foreign value.

**Notes**

Olivier Cadot is Professor of Economics at the University of Lausanne, associated scholar at Centre d'Etudes et de Recherches sur le Développement International (CERDI), and fellow at the Centre for Economic Policy Research (CEPR); his email address is olivier.cadot@unil.ch. Jaime de Melo (corresponding author) is Professor of Economics at the University of Geneva, associated scholar at CERDI, and fellow at CEPR; his email address is demelo@ecopo.unige.ch. The authors thank Paul Brenton and Marcelo Olarreaga for many useful suggestions and their colleagues and co-authors Céline Carrère, Antoni Estevadeordal, Alberto Portugal-Perez, Akiko Suwa-Eisenmann, and Bolormaa Tumurchudur for permission to draw on joint work. They also thank three referees for comments on a previous draft.

1. According to this same tally, 45 developing economies having signed bilateral trade agreements with a developed country, and 90 of the 109 preferential trade agreements between developed and developing economies have been created since 1990.

2. According to a survey administered by the World Customs Organization to customs officials in developing economies (as reported by Brenton and Imawaga 2004), 67 percent of respondents in...
Sub-Saharan Africa agree that dealing with rules of origin under overlapping trade agreements causes problems, and a majority also agrees that rules of origin are more labor-intensive. Administering rules of origin detracts from other objectives of tax collection and trade facilitation.

3. Because meeting the requirements is difficult and appears unnecessarily complex, in view of the European Commission's objective to grant some preferential access to its market for GSP-eligible countries, on 16 March 2004 the commission adopted Communication COM (2005) on "The Rules of Origin in Preferential Trade Arrangements." The communication explores alternative rules of origin that would be simpler and more development friendly. A key proposal under consideration is to replace the current product-specific rules of origin with a single rule based on a minimum of originating value added.

4. By comparison, the average preferential margin (computed over tariff lines with positive tariffs) was 4.5 percent for NAFTA (almost all tariffs had been eliminated on NAFTA trade by 2001), 2.4 percent for GSP-eligible countries, and 4.6 for ACP countries (not eligible for Everything But Arms status). Data for the European Union are for 2004, when 62 percent of trade for GSP-eligible countries and over 80 percent of trade for ACP countries took place at zero tariffs (some ACP also benefited from Everything But Arms status at zero tariffs in the EU market).

5. Krishna (2006) discusses other effects that are more difficult to quantify: effects such as rules of origin-jumping investment and effects on intermediate prices. Thoenig and Verdier (2006) also consider the implications of rules of origin for multinationals confronted with outward-processing decisions.

6. The United States rarely uses a value content criterion as the sole requirement for origin, and when it does it tends to rely on a single 40 percent foreign content requirement. The European Union has value content criteria ranging from 50 to 15 percent of domestic value added.

7. Cumulation is, in principle, only diagonal (see the glossary in the appendix), but the domestic content can be calculated as an aggregate of value added in any ASEAN member state; so in effect AFTA provides for full cumulation, although, as noted by Brenton (2006), the rules stipulate that the final stage of manufacture must be carried out in the exporting member state (what constitutes "the final stage" is not defined). Because vertical links and outsourcing are very important in Asia, full cumulation considerably relaxes the requirements of satisfying origin.

8. To drive home the importance of trade in intermediate goods, consider the following example. On the basis of the input–output data in Baldwin (2006, table 1 for Indonesia, Malaysia, Philippines, and Thailand (middle-income Asian countries), an average of 35–40 percent of intermediate goods are sourced outside AFTA. For example, take an activity with 10 percent value-added and 40 percent nonoriginating intermediate goods—that is, 36 percent of the final unit product price is nonoriginating. Originating value for this activity would be 64 percent. Then take the plausible example of an activity with the same value added but with 60 percent of materials nonoriginating; originating value falls to 46 percent, barely above the 40 percent minimum currently stipulated in AFTA.

9. The special regime was recently extended until 2015. Figure 2 lists the 22 beneficiary countries.

10. The algorithm is in essence a grid search over cutoffs whose criterion is the minimization of the concentrated sum of squared errors of the ordinary least squares regressions in the two regimes.

11. The decline in exports to the preferential-giving destination suggests that producers would choose to export under most favored nation status. In the illustrative simulations reported here, with constant elasticity throughout and smooth substitution possibilities across the origin for intermediate good purchases and export destination sales, producers pass on cost increases to consumers.

12. The authors of this article are aware of concerns voiced by the private sector in the course of the EU review about the practical difficulty of a value content criterion for small firms and, if based on costs, its potential to force unwanted disclosure of strategic information to powerful EU buyers that would enhance their ability to squeeze rents from developing country producers.
References


Cadot, O., C. Carrère, and V. Strauss-Kahn. 2007. "Export Diversification: What’s Behind the Hump?" University of Lausanne, Switzerland.


Oxford Journals Advertising

For information on advertising in this journal please contact:

Helen Pearson
Oxford Journals Advertising
PO Box 347
Abingdon, OX14 1GJ, UK

Tel: +44 (0) 1235 201904
Fax: +44 (0) 8704 296864

e-mail: helen@oxfordads.com

Information available online
Visit www.oxfordads.com
THE WORLD BANK

1818 H Street NW
Washington, DC 20433, USA
E-mail: researchserver@worldbank.org