Conceptualizing and Measuring Institutions: A View from Political Science

Marcus Kurtz
Department of Political Science
Ohio State University
kurtz.61@osu.edu

Andrew Schrank
Department of Sociology
University of New Mexico
schrank@unm.edu

Initial conference draft: Neither for circulation nor for citation

Introduction

The following paper identifies and takes issue with a potentially troubling development in the study of comparative politics. While political scientists have traditionally deployed *objective* indicators of "political organization and administrative capacity" (Reynolds 1983, p. 976), including tax ratios (Organski and Kugler 1980; Benson and Kugler 1998), tax structures (Kling 1968; Krasner 1985), political participation (Przeworski and Sprague 1971), and the nature and extent of public service provision (Migdal 1988; Putnam 1992), they are beginning to import *subjective* indicators of "governance" from economics (Sandholtz and Gray 2003; Fish 2005; Gerring and Thacker 2005; Blake and Martin 2006; Cameron et al. 2006). In fact, the customary distinction between objective and subjective measurement is by now less salient than the various differences *among* the subjective or "perceptions-based" alternatives (Kaufmann et al 2005a). Whose perceptions do they capture? Of which institutions or issue areas? And to what effect? We hold that the leading perceptions-based indicators are poorly explicated; give the perceptions and interests of businesspeople (and foreign businesspeople in particular) undue influence; are riddled with measurement error and all but impossible to interpret; and should therefore be either reconsidered or replaced by *improved* objective indicators that take public sector inputs as well as outputs into account. We make the case for reconsideration and/or replacement in four principal sections. Section 1 offers an abbreviated introduction to the perceptions-based literature. Section 2 develops a critique of the conceptual and operational underpinnings of the leading perceptions-based indicators. Section 3 discusses the problem of random measurement error in practice. And Section 4 discusses objective alternatives.

1. Intellectual context.

Economists first brought perceptions-based indicators of corruption, property rights, and the rule of law to bear in cross-national growth research in the mid-1990s. For instance, Paulo Mauro obtained national-level indicators of judicial integrity, corruption, and bureaucratic quality from commercial risk rating agencies as early as 1995 (Mauro 1995). Arthur Goldsmith deployed an index of property rights developed by the Heritage Foundation more or less concurrently and turned to the Corruption Perceptions Index developed by Transparency International for the first time a few years later (Goldsmith 1995, 1999). And Alberto Ades and Rafael DiTella introduced the corruption indicators found in the EMF Foundation's *World Competitiveness Report* at approximately the same time (Ades and DiTella 1997).

By the end of the decade, however, Daniel Kaufmann and his colleagues at the World Bank Institute had come to believe that "many of these indicators serve as imperfect proxies for one of a smaller number of more fundamental concepts of governance," including "the rule of law, government effectiveness, and graft" (Kaufmann et al. 1999a, pp. 1-2), and had therefore decided to develop a series of "aggregate governance indicators" designed to capture the fundamental concepts in a more transparent and accurate fashion. They aggregated data from a wide array of sources including—but by no means limited to—the aforementioned foundations, nongovernmental organizations, and commercial risk rating agencies;

developed an unobserved components model; and distilled the aggregate data into six different indicators: voice and accountability, political instability and violence, government effectiveness, regulatory burden, rule of law, and control of corruption.

The so-called World Governance Indicators (WGI) have gained rave reviews (Gervasoni 2006) and market share in political science, sociology, and law, public policy, and of course economics (see, e.g., Fish 2005; Rigobon and Rodrik 2005; Apodaca 2006; Borrrmann et al. 2006; Lee 2007)—in no small part due to the creativity and entrepreneurship of the authors. And the WGI are by no means devoid of merit. They are available for more than 150 countries. They cover a ten year period. They include standard errors as well as individual country scores. And their authors issue frequent and welcome caveats as to their various limitations (see, e.g., Kaufmann et al. 2005b, p. 41).

In fact, the WGI arguably provide the best possible case for the perceptions-based measurement of governance. They are by all accounts among "the most carefully constructed and widely used" (Arndt and Oman 2006, p. 49) indicators available. And they are by now more than a decade old. If they fail to convince, we argue, then the entire project of perceptions-based measurement demands reconsideration—if not abandonment.

2. The limits to perceptions-based indicators in prospect: from explication to operationalization.

Do the WGI convince? We are neither the first nor the only scholars to entertain doubts (Arndt and Oman 2006; Thomas 2007; Knack 2007; Bhagwati 2007). Our doubts begin, however, with the explication of the conceptual underpinnings of the measures themselves (see Kurtz and Schrank 2007a, 2007b). Rudolf Carnap famously defined the process of explication as "the transformation of an inexact, prescientific concept, the *explicandum*, into a new exact concept, the *explicatum*" (Carnap 1962, p. 3), and offered a number of apposite examples including the self-conscious transformation of the *explicandum* of "warmth" into the *explicatum* of "temperature." According to Carnap, the description of the *explicandum* is a necessary starting point in the process of measurement and comparison and is therefore no less central to scientific progress than the eventual interpretation and assessment of the *explicatum*. "Although the *explicandum* cannot be given in precise terms," he argued, "it should be made as clear as possible by way of informal explanations and examples" (Carnap 1962, p. 3). Otherwise parties to a scientific discussion or dispute would inevitably wind up talking past one another.[1]

Our discussion therefore departs from the *explicanda*. What are the prescientific concepts in question? While Kaufmann and his colleagues purport to be measuring institutions, and devote a good deal of time to their operationalization, they offer ad hoc and/or inconsistent definitions of their underlying concepts. Take, for example, their measure of the "rule of law." They initially define the rule of law as "respect of citizens and the state for the rules which govern their interactions" (Kaufmann et al. 1999b, p. 2). They subsequently abandon their original definition for "the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence" (Kaufmann et al. 2005a, p. 4). And they eventually decide to deploy a more encompassing, if less discriminating, definition that includes the "extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence" (Kaufmann et al. 2006, p. 4). They offer neither Carnap's "informal explanations and examples" nor for that matter a rationale for their definitional amendments—let alone a discussion of their implications. Are we to assume that their rule of law indicator (RL) assesses different concepts in different years? Or that

[1] Dumont and Wilson import Carnap's approach to explication into the social sciences (1967). Przeworski and Sprague (1971, p. 217) embrace Hempel's extension of Carnap's approach. And Wesley Salmon underscores the continued relevance—if by no means immutability—of the broader realist program to which Carnap and Hempel contributed (1999).

the definitional changes have no effect on the indicator's interpretation? And, if the latter, why do they introduce the changes in the first place?

The implications are by no means trivial. Let's take a concrete example. Italy's rule of law score all but collapsed between 1996, when the Italians (.97) outperformed the Greeks (.90), Czechs (.84), Slovenes (.86), and Koreans (.71), and 2006, when their estimated governance score fell to .37 with a standard error of .14 and they suddenly found themselves relegated to an ignominious position beneath Tunisia (.38), Jordan (.45), Botswana (.63), and Uruguay (.40)—not to mention a number of monarchies, dictatorships, emirates, and tax havens (Kaufmann et al. 2007b, Table C5). In fact, Italy's estimated RL scores have declined from their previous values in seven out of eight releases of the WGI. Why? Have Italian politics and society really change so dramatically and so consistently over the course of a single decade? Or did Kaufmann and his colleagues de-emphasize the informal relationships and institutions that loom large in the Italian political economy (Putnam 1992) in favor of formal institutions like the police and courts—that are less operative in much of Italy—when they abandoned their original definition of the rule of law? Our point is neither to whitewash Italy's very real governance dilemmas nor to rule out other explanations for the fall. We simply wish to underscore the fact that conceptual shifts—especially those that are neither explained nor justified by their authors—come with real interpretive costs.

Our concerns are neither frivolous nor pedantic. After all, the rule of law's empirical referent is anything but obvious. Legal scholars like Thom Ringer have not only condemned the "conceptual anarchy among development theorists, experts, and donor agencies surrounding the very meaning of the expression" but have gone on to wonder whether the "problem with measuring the success of rule of law reform initiatives is that the parties assessing them may have something quite different in mind to those implementing them" (Ringer 207, p. 182; see also Daniels and Trebilcock 2004). Nor are they alone. Political scientists and sociologists have expressed doubts about the concept as well (Stephenson 2000; Kurtz and Schrank 2007b). And Dani Rodrik has recently wondered whether he's "the only economist guilty of using the term abundantly without having a good fix on what it really means" or simply "the first one to confess to it" (Rodrik 2007).

Kaufmann and his colleagues appear to be sanguine by way of comparison. While they take comfort in the fact that the various aspects of governance they purport to measure "tend to be quite highly correlated across countries" (Kaufmann et al. 2007a, p. 555), they are in fact assuming what needs to be proven—that is, that their correlated indicators are valid measures of governance in the first place. Otherwise the reported correlations are all but devoid of meaning (see also Knack 2007).[2]

Are the WGI valid indicators of governance? We will address the question by focusing primarily upon the most recent iteration of their rule of law indicator, for "measurement validity is specifically concerned with whether operationalization and the scoring of cases adequately reflect the concept *the researcher* seeks to measure" (Adcock and Collier 2001, p. 529; our emphasis), and Kaufmann and his colleagues

[2] In fact, Kaufmann and his colleagues are at least implicitly adopting the by now discredited "operationist" approach to measurement in which meanings are produced by measures and not vice versa. While they explicitly claim to derive the definitions of the six aspects of governance covered by the WGI from "existing definitions or understandings of the concepts" (Kaufmann et al. 2007c, p. 24), they occasionally admit that their indicators actually drive their definitions—and that the meaning is therefore in the measure. "That is," they write, "we have just one implicit definition of corruption, which comes from the aggregation of these many data sources across many countries" (Kaufmann et al. 2007c, p. 7). The limits to operationism are by now well known. According to Henry Byerly the "strict application of the more extreme operationist doctrine would lead to as many notions of a quantity such as mass as there are different operational procedures for measuring mass" (Byerly 1972, p. 376.). Less strict applications are untenable, however, for efforts to justify—rather than simply invoke—particular operational definitions inevitably appeal to meanings that are independent of the measures themselves. See Hempel (1956) for a seminal critique of operationism.

hold that RL in particular captures the "norms of limited government" (Kaufmann et al. 2007a, p. 555) that are central to the literature on institutions and growth (Kaufmann et al. 2007a, p. 561; see also Kaufmann and Kraay 2002, p. 192).[3]

Table 1 recapitulates the definition of the rule of law that appears in the most recent iteration of the WGI—i.e., the concept that Kaufmann et al. purport to measure—and includes summary information on the sources, respondents, and questions incorporated into the RL indicator. The questions tend to fall into three principal subgroups: crime, property rights, and judicial and security institutions. We evaluate RL's content validity by reflecting upon each subgroup and asking (i) whether key elements are omitted or (ii) inappropriate elements are included in the indicator's construction (Adcock and Collier 2001, p. 538).[4]

PLEASE INSERT TABLE 1 ABOUT HERE

Crime. RL includes data on both crime in general and a number of specific offenses (i.e., kidnapping) or categories of crime (i.e., organized crime). The underlying data sources include expert assessments of crime prevalence rates and trends, household survey data on victimization, and subjective assessments of the costs that criminals impose on business. And the rationale for their inclusion is more or less obvious: A society that is marked by the rule of law is unlikely to be crime-ridden and vice versa.

Our concerns therefore derive not from the incorporation of crime data *per se* but from the manner in which they're incorporated. After all, the data summarized in Table 1 are unlikely to present an unbiased portrait of the level and distribution of crime in society. On the contrary, they are likely to give undue weight to crimes that affect business and elite interests. While the surveys undertaken by the World Bank and the World Economic Forum are self-consciously designed to tap the costs crime imposes on business (see, e.g., Kaufmann et al. 2007b, Table B5), the household surveys undertaken by Gallup and the various regional barometers are likely to prioritize the interests of the elite for a less obvious reason. Respondents in crime victimization surveys are known to overreport property crimes that tend to affect the better off and underreport violent crimes that tend to affect their less fortunate compatriots (Bergman 2006, p. 221).

Thus, the crime data incorporated into RL are best portrayed neither as proxies for the rule of law writ large nor as random variables but as indicators of the elite's ability to insulate itself from what Alejandro Portes and Brian Roberts have labeled the "forced entrepreneurship" (2005, p. 67) of the poor in the Latin American context. According to Portes and Roberts, the relentless onslaught of debt, austerity, and free market reform has provoked the unprecedented growth not only of poverty and inequality but of anger and resentment in Latin American cities. "Property crime may rise in these contexts," they maintain, "as

[3] Adcock and Collier worry that scholars who fail to put their conceptual difference aside for the purposes of measurement validation will find themselves paralyzed by intractable disputes over concepts (2001, pp. 538-9). We therefore follow their advice and take Kaufmann et al.'s definition of the rule of law as a given for the purpose of the discussion of validity—despite our skepticism about their conceptualization more generally. We should note briefly that we find the conceptualization—but not the operationalization—lying behind their "government effectiveness" measure to more adequately capture the classical Weberian notion of good governance. See Kurtz and Schrank (2007a, 2007b) for a discussion.

[4] An alternative approach would assess convergent and discriminant validity by asking whether RL (i) is strongly correlated with different measures of the rule of law and (ii) weakly correlated with indicators of distinct but related governance concepts. While Kaufmann and his colleagues tend to *incorporate* the available indicators of rule of law into their own metaindicator, and thereby render RL more or less immune to convergent validation, they boast that RL is *strongly* correlated with their other governance indicators (Kaufmann et al. 2007a, p. 555)—and thereby raise serious doubts about their divergent validity.

some members of the subordinate classes take matters into their own hands in order to redress both absolute and relative deprivation."[5]

Our point is most assuredly neither to rationalize property crimes nor to deny their relationship to the rule of law but to underscore the fact that they are neither (i) inclusive of the full array of crime in most societies nor (ii) random occurrences and that RL therefore *omits* important elements of the rule of law (i.e., violent crime, violations of labor and environmental law, etc.) and inappropriately—if indirectly—*includes* the various correlates of crime at the country level (e.g., poverty, inequality, the age distribution of the population, collective efficacy, etc. [see, e.g., Sampson et al. 2007]).

Property rights. RL includes data on property rights and threats to their sanctity (i.e., expropriation) as well as information on a number of specific categories of property right (i.e., intellectual property). The property rights indicators are drawn from commercial risk rating agencies, firm-level surveys, and inter- and nongovernmental organizations. But the rationale for their inclusion is unclear, for Kaufmann et al.'s definition of the rule of law neither makes specific reference to property rights nor militates against their compromise or outright transgression by *legal* means.

Take, for example, eminent domain provisions in United States property law. They give the government the authority to override legitimate property claims in a host of well-defined situations, and are therefore perfectly consistent with the rule of law, but nonetheless give rise to virulent opposition in practice, and are therefore likely to undercut the RL score if used too frequently.

Our objection is readily addressed, however, neither by asserting the infrequency of eminent domain nor by retroactively including a reference to property rights in the definition of RL, for the expropriation and confiscation of property are frequently deployed in the war *against* corruption and crime and their use is therefore no less consistent with the growth than the absence of the rule of law. Park Chung Hee's campaign against the "illicit accumulators" in South Korea offers a case in point (Schrank 2007), and more recent examples are readily available in Africa, Asia, and Latin America (see, e.g., Economist 1981; Arnold 1999).

But one need not travel to the developing world to witness the confiscation and expropriation of assets deployed in defense, rather than contravention, of the rule of law. The Racketeer Influenced and Corrupt Organizations (RICO) Act offers a no less apposite North American example. After all, RICO gives US government prosecutors the right to seize the assets of allegedly corrupt or criminal organizations before they have attained a conviction—and thereby threatens the economic lifeblood of suspect enterprises. A number of observers have given RICO credit for the demise of the Mafia in the 1990s (Laurence 2004). And James Jacobs and Lauryn Gouldin have therefore labeled the act "the most important substantive antiorganized crime statute in history" (Jacobs and Gouldin 1999, p. 169).

The statute has not, however, been free of criticism. While law enforcement officials are justifiably fond of RICO, and favorably disposed toward asset forfeiture in particular, their critics decry "policing for profit" (Blumenson and Nilson 1998) and worry that the law not only gives the government a "license to steal" (Levy 1996) from legitimate businesses but constitutes a "nuclear deterrent to rational negotiations" as well (Arkin in Newsweek 1989). Is the cure of RICO worse than the disease of corruption and crime? The answer is anything but obvious, for the RICO experience suggests not only that the same institution

[5] Our own admittedly crude data analysis confirms their sense that Latin America's current crime wave betrays a pattern of "entrepreneurial rationality" (Portes and Roberts 2005, p. 75). We regressed crime victimization data from Latinobarometer (1 = victim within the past year) on a seven point scale designed to capture the respondent's level of education—an admittedly crude proxy for social class—in 2005 and found an enormous positive effect (odds ratio 1.39; p value < .0001). Results available from the authors upon request.

can simultaneously militate in favor and against the rule of law but that property claims themselves are by their very nature subjective and controversial—even in the advanced industrial countries.

They are arguably more controversial in the developing world, however, and are therefore decidedly ill-equipped to play a meaningful role in perceptions-based indicators of the rule of law. "A government that evicts squatters will in all likelihood be portrayed as a threat to private property by the squatters and a bulwark against expropriation by the landlords," we have argued, "and the problem is likely to be compounded by the fact that in much of the developing world this year's squatters are likely to be next year's landlords and vice versa" (Kurtz and Schrank 2007b).

Nor are the examples of allegedly ill-gotten gains in the US and squatters in the developing world unique. The problem of ambiguous property claims is widespread and will in all likelihood be aggravated by the growing salience of intangible or intellectual property in the future (see, e.g., Evans 1997), for different states and societies have radically different conceptions of the appropriate length and scope of patent, copyright, and trademark protection.

Judicial and security institutions. RL also includes questions on courts, contract enforcement, and the police. The "courts, cops, and contracts" cluster, as we call it, incorporates data from firm and household surveys as well as expert assessments; tracks the government's ability and willingness to ensure the safety and security of the population; and plays an almost indisputable part in the assessment of the rule of law.

We nonetheless wonder why Kaufmann and his colleagues simultaneously omit or de-emphasize a series of regulatory agencies that are no less central to the security and well-being of the population including but by no means limited to the tax, labor, food, drug, health, and safety inspectorates; banking and financial overseers; and environmental monitors. Regulatory authorities are involved in *law enforcement*, after all, and their virtual exclusion from RL arguably says more about the types of laws valued by the indicator's authors and advocates than about their importance to the concept itself.[6]

In short, RL appears to suffer from a pronounced and systematic pro-business bias. The crimes covered, the institutions included, and the interests served all point in the same direction—as do the interests, institutions, and crimes overlooked. And the biases will arguably prove not only threatening but fatal to content validity when aggravated by what for lack of a better term we'll call sampling error. After all, the bulk of the data incorporated into the measure—approximately two-thirds of the "representative" sources that are weighted most heavily in the final indicator (see Table 1)—are derived from either private investors or their commercial advisers.[7]

Kaufmann and his colleagues admit that pro-business biases are possible in theory but belittle their impact in practice (2007a, p. 556). We've already established that reported crime victimization varies by social class. Do business perceptions of other aspects of the rule of law differ substantially as well? Fortunately, the Latinobarometer data employed by Kaufmann and his colleagues invite a direct test.

PLEASE INSERT TABLE 2 ABOUT HERE

Table 2 includes the results of two ordered logistic regression models. The dependent variables are the answers to the questions about trust in the judiciary and police that are incorporated into the RL indicator.

[6] A brief review of the WGI "regulatory quality" (RQ) indicator does little to assuage our concerns. RQ addresses the degree to which tax, labor and environmental laws compromise business competitiveness, for example, but fails to ask to what degree they protect workers or the environment (Kaufmann et al. 2007, Table B4).

[7] Another comes from a nongovernmental organization that the authors themselves describe as "conservative." And only one source includes any citizen surveys at all.

They are coded from 1 (much trust) to 4 (no trust). They are regressed on an indicator variable that assumes the value of 1 wherever the respondent is a businessperson and a series of country dummies. And they underscore the differences between businesspersons who consistently evince less faith in police and judicial institutions than their compatriots.[8]

A defender of the WGI might rebut charges of bias by portraying businesspersons and their advisers as particularly desirable data sources whose opinions should loom large in their indicators. Private investors not only have more experience with institutions like the police and courts than the average citizen, they might argue, but are also in the vanguard of the struggle for growth and development in market societies. But non-businesspersons are neither ignorant nor unimportant to the process of growth and development; on the contrary, they make countless individual decisions—about everything from personal savings and voting to the treatment and education of their children—that aggregate into important societal outcomes. Such decisions not only influence but are influenced by the broader socio-political context, however, and a governance indicator worthy of the name will therefore take the ideas and interests of non-businesspersons seriously.

In short, we have argued not only that the concept of rule of law is poorly explicated but that the RL indicator itself is invalidated by skewed questions addressed toward a biased sample of respondents. The consequences are particularly acute when RL is incorporated into growth regressions as an exogenous variable, for a positive coefficient could reflect any combination of at least four different underlying processes: first, a positive relationship between *actual rule of law* (e.g., crime control, confidence in the courts) and growth; second, a positive relationship between *some but not all aspects of the rule of law* and growth; third, a positive relationship between the *underpinnings of the rule of law* (e.g., equity, opportunity, etc.) and growth; or, fourth, *herd behavior* on the part of investors who receive their advice from the same consultants and risk rating agencies and thereby animate growth regardless of the so-called fundamentals.

3. The limits to perceptions-based measures in practice: measurement error.

So far we have emphasized the problem of non-random measurement error (systematic bias) in the WGI's preferred RL indicator.[9] While systematic errors are in all likelihood pervasive not only in RL but in other perceptions-based governance measures, they are by no means the only worry. We are also concerned about random measurement error—pure noise—in analyses that rely on subjective indicators.

Kaufmann and his colleagues admit that their indicators are *estimated* quantities and are therefore marked by at least some amount of random error. The indicators are estimates of the mean of a distribution rather than population parameters and as such they come with standard errors attached. In fact, Kaufmann et al. are refreshingly candid about the estimated nature of their indicators. Many of the measures used by political scientists including democracy scores, poverty rates, national socio-economic characteristics derived from surveys, and even election vote shares are estimated quantities—though this is rarely acknowledged.

[8] Kaufmann et al. appear to believe that systematic differences between business and non-business perceptions can only contaminate their estimates if they alter the rank order of countries (2007a, p. 570). We can neither confirm nor rule out alterations in rank order given the complete absence of citizen surveys for most of their indicators but we are not convinced that they are necessary to introduce bias into RL—which is, after all, an interval measure. Furthermore, the Afrobarometer data reported in Schrank and Kurtz (2007b, p. 566, Table 1) suggest that businesspeople themselves offer inconsistent evidence on the quality of governance. While they actually report better access to public services than their compatriots, they nonetheless hold the government in lower esteem.

[9] It is difficult, of course, to quantify the size of such biases without some means of estimating the 'true' level of the rule of law.

We maintain that problems of *random* measurement error are no less threatening to the WGI than the systematic bias introduced through the incorporation of subjective and selective indicators. Our discussion departs from the realization that the admittedly appealing country-year specific standard errors found in the WGI are likely to be low-end estimates of the actual random variation. We move on to the fact that the estimates of measurement error are—for any number of perfectly reasonable reasons—correlated with other features of the data including, but not necessarily limited to, the level of per capita income. Since the imprecision of the estimates varies in systematic ways, this has implication for the weights each case has in any analysis that properly accounts for it. In a regression framework, for example, this would necessarily give wealthier countries with better educated populations (which are characterized by better-measured governance) more weight in the calculation of the parameter estimates. This is not necessarily a problem unless assumptions about causal homogeneity in the data are not fully met. Finally, we examine the consequences of measurement error for causal analysis through an errors-in-variables framework and directly explore the implications of imprecision under the assumption that the WGI's reported standard errors are in fact accurate, and alternatively, if they were to be greater than their reported values. When such errors are accounted for in multivariate analyses, we argue, the effects can be quite substantial—not only for the effect estimates of governance but also for estimates of the effects of other variables of interest – even where they are measured without error. Ultimately, this further reduces our confidence in such subjective measures and underscores the need for an alternative strategy for measuring governance.

What's the Problem with (Random) Measurement Error? The problem of measurement error as it applies to the use of the WGI—or any other data—takes two forms. In a regression framework where a variable measured with error is the dependent variable, it is commonly assumed that aside from a loss of efficiency, there is little problem—for the uncertainty is essentially absorbed into the error term of the regression. As Lewis and Linzer (2005, 346) have pointed out, however, this is true only where the "sampling uncertainty is constant across observations," and where it is not, the regression errors will be heteroskedastic and OLS estimates will be inconsistent.[10] This is not, however, our principal concern. Rather, we worry about the effects of measurement error when governance indicators are independent variables in a regression analysis. As is well known, in a multivariate regression, measurement error in the independent variable causes bias in the parameter estimate for that variable, as well as having effects on any other independent variables with which it is correlated.

This is, of course, the well-known errors-in-variables problem. The standard framework in which this is explored assumes a data generating process that involves a known (and constant) amount of uncertainty for each observation.[11] This is not the case with the WGI, where the standard errors vary quite a bit from observation to observation. Our approach will be to directly incorporate this information in a simulation-based effort to understand the effects of this uncertainty in several straightforward examples. The point is emphatically not to provide anything like a definitive model of the relationship between governance and (in this case) growth, but rather to demonstrate that even taking for granted the reported standard errors of the WGI data, measurement error poses a substantial problem for inference.

WGI errors are probably too low. The WGI employs a strategy of aggregation in constructing its governance indicators that should, in principle, minimize the amount of measurement error. The idea is to use as many sources of data as possible and to weight them by the extent to which they produce similar

[10] Indeed, standard solutions are not necessarily effective either, as Lewis and Linzer (2005) point out, the errors of the regression have two parts—that coming from the mismeasurement of the dependent variable, and that which would be present even were measurement perfect. If the latter portion were homoskedastic then standard remedies for heteroskedasticity caused by the former would also produce inefficient and potentially inconsistent results.

[11] For example, this would be the case with a vote-counting machine that has a known failure rate. Each voter confronts an identical probability that her vote will be misrecorded.

results. If they are unbiased and independent, then this strategy should reduce overall measurement error. If the inputs are biased, however, the weighting strategy becomes problematic. But what is of concern here is the independence of the sources.

As Knack (2007, 266-268) has shown, quite a lot of the sources treated as independent in constructing the WGI in fact rely on each other to come up with their individual ratings.[12] Indeed, this problem is one of the reasons we contend in the concluding section of our paper that an approach that relies on objective measures—rather than subjective and non-independent assessments—will ultimately produce more useful governance metrics. And in this context, if sources are not fully independent of each other, the standard errors reported alongside governance estimates are likely to be too low—for they are calculated on the assumption that there is more independent information available than is actually the case. For this reason, while we will continue to rely on these estimates in the analyses that follow, we wish to emphasize that these are inevitably low-end estimates of the problems that may be present, and that the real biases induced by measurement error are likely (substantially) larger even than those that we report below. To test the implications of potentially larger true standard errors for these estimates, we have examined parallel analyses under the alternative assumption that the true errors are 50 percent and 100 percent greater than those reported in the WGI.

How reliable are WGI measures? Our first task is to examine how big a problem purely random measurement error is for the WGI. In this and subsequent examples we will continue to focus on the RL indicator though the method is appropriate for the examination of any such measure. To begin, we ask whether repeated observations—each incorporating an element of randomness defined by the standard errors reported by KKM that apply to each country-year—are highly correlated with each other. Table 3 reports the correlations among repeated observations of the 1996 rule of law measure under three conditions: (1) assuming the standard errors are as reported by KKM, (2) assuming the standard errors are 50 percent greater than reported, and (3) assuming they are 100 percent larger than reported.

Please insert Table 3 about here

The estimates and confidence intervals reported in Table 1 represent the median of 1000 simulations (for the correlations) and the 2.5th and 97.5th percentiles, respectively, for the upper and lower bounds on the 95 percent confidence interval. We see that, if the KKM standard errors are in fact accurate, a moderate degree of reliability is obtained by the measure. The median correlation among repeated observations is 0.909, with a relatively narrow confidence interval. But if, as we have argued above, the real underlying measurement error is somewhat higher, there is greater cause for concern. For example, if the standard errors of the RL estimates are even 50 percent greater than reported, the median correlation between repeated observations of this measure falls to 0.818, with a wider confidence interval (0.752–0.867). If the actual measurement error is twice as severe as estimated by KKM, then reliability is even more profoundly affected, with repeated observations correlating at only 0.716, with an even wider confidence interval (0.607–0.788). While these correlations may seem relatively strong, it is important to keep in mind that they represent repeated observations of *identical* phenomena, and thus one would expect them to be highly correlated. It is also the case, when these are simply used as data in a regression model, it is *assumed* that the correlation among repeated observations is perfect—i.e., that the variable is measured without error. In the next section, we explore the consequences of the failure of this assumption.

It is also important to note—for entirely understandable reasons of data availability—that the estimated standard errors for the RL measure are negatively correlated with the income per capita of the country in question. Better data (typically, more sources) are available for wealthier countries. But this has the possibility to wreak havoc on efforts to use these measures in analyses of developing countries (and it is

[12] Give some examples here.

in precisely these contexts that the WGI are most commonly used). The second panel of Table 1 replicates the analysis above, in this case restricting the sample to countries below the median national income. But in this context, the reliability of the RL measure in 1996 is far worse. Even if the reported uncertainty in the KKM measures is correct, for the poorer half of the globe the median correlation among repeated observations is only 0.724, and the confidence interval expands (0.581 – 0.821). But if these errors are actually larger, reliability plummets, to a median correlation of 0.538 where the errors are half again as large, and to 0.398 if they are double the reported estimates. At a minimum, this raises concerns about the discriminating power of the rule of law indicator, at least for developing countries. And as one can see from the confidence intervals (Table 1) associated with these estimates, we have very little certainty about these reliability estimates.

What do we learn by incorporating estimates of uncertainty into analyses? We wish to be very clear that it is a signal feature of the WGI indicators that they include with them explicit estimates of the uncertainty that surround each point estimate. Indeed, in cross-national data efforts the provision of indicators of uncertainty is rare indeed—and unrelated to the reliability of the measures themselves. Indeed, recent efforts to calculate such information for the widely-used Polity measures of democracy suggest that measurement uncertainty is quite substantial, often enough to undermine confidence in well-established results relying on them (see Treier and Jackman 2006). What we examine below are the consequences of measurement error for analyses that rely on the WGI measure of the rule of law—both for the effect that this has on the coefficient estimates for this variable but also its implications for other included independent variables.

In the analysis that follows we examine the widely-expected relationship between the rule-of-law and economic development. We do not pretend that the cross-country growth regressions presented below represent anything like the state of the art in such analyses; our task is quite different. We are interested in the effects that measurement error has on coefficient estimates in straightforward models of economic growth. And we expect that if it poses substantial challenges in such a simple OLS regression framework, these may be equally (if not more) severe in other contexts.[13]

We estimate a simple prospective model, as we believe it a reasonable approach to the endogeneity that bedevils cross-national growth modeling. We are interested in the forward-looking effects on economic growth of variations in the rule of law. Since the earliest measurement available to us is from 1996, this is our input. We then construct a dependent variable—the average annual rate of growth over the 1997-2004 period—that minimizes the effects of year-on-year volatility in economic growth, but looks only at economic growth subsequent to the rule of law measurement (and thus is free of any potential halo effects). In addition, we include controls for the level of investment, the secondary enrollment rate, the natural log of GDP/capita, the log of population, and a set of regional dummies to (at least partially) capture unmeasured heterogeneity. These are all measured for 1996 or before, to minimize endogeneity. The full equation is as follows:

$$\text{Growth}_{1997\text{-}2004} = \text{Rule of Law} + \ln(\text{GDP/cap}) + \text{Secondary Enrollment Rate} + \text{Investment} + \text{Europe} + \text{Middle East} + \text{Africa} + \text{Latin America} + \text{Asia} + \varepsilon$$

We employ a simulation-based approach to estimate the consequences of measurement uncertainty in the rule of law measure. Thus, for each of a 1000 simulations, we "draw" a rule of law score for each country taken from a distribution with a mean set to the WGI rule of law estimate and the appropriate standard deviation associated with that mean. Each of these simulations generates a set of coefficient estimates for the variables in the model, the median of which forms our final estimate of the coefficients, while the 2.5th

[13] For example, Yatchew and Griliches (1985) have shown that measurement error is even more problematic in the probit context.

and 97.5th percentiles in this distribution mark out the empirical 95 percent confidence interval. For exploratory purposes we also performed these simulations under the assumption that the level of measurement variability is 50% and 100% greater than suggested in the WGI.

Figure 1 graphically reports the results of these simulations for the coefficient estimates on the rule of law. Green dotted lines represent the 95% confidence intervals for each individual regression comprising the simulations, while black dashed lines mark off the 2.5th and 97.5th percentiles of the distribution of simulations. The degree of uncertainty induced by measurement error can be discerned from the deviation of each of the curves from a perfectly horizontal line. Thus we see that, where we assume that the standard errors reported by Kaufmann et al. are correct, the estimated coefficient on the rule of law is -0.325, but with a confidence interval that spans [-0.68, 0.062]; since this interval includes zero, the coefficient is statistically indistinct from zero at conventional levels of significance. Interestingly, as the level of uncertainty increases, the coefficient estimate is attenuated—it becomes closer to zero. At the same time, naturally, the 95% confidence interval around it increases. The coefficient estimates for rule of law using the reported standard errors are ***larger*** in absolute value than those produced in a simple OLS regression obtained under the assumption that the data are measured without error (in this case, β_{OLS} = -.197, compared to $\beta_{simulated}$ = -0.325).

Nor are the biases necessarily confined to the estimated effect of the rule of law. While the coefficients on Secondary Enrollment and Investment are largely unaffected by the uncertainty in the rule of law measure, the level of development (measured as ln(GDP/capita) is not. In Figure 2, we report the effects on estimates of the effect of the level of development in the same model that come as the result of measurement errors in the rule of law indicator. Notably, as the amount of measurement error in the rule of law indicator increases, the estimated coefficient on the level of development ***increases*** in absolute value—underscoring the point that the direction of the bias induced by measurement error is not ex ante predictable (save in the bivariate case).

We were most concerned, however, with the problem of measurement error as it affected the less developed countries. Thus, in Figure 3 we present a replication of the analysis we carried out for Figure 1, but in this case confining the dataset to countries at or below the median level of GDP/capita. Here we see that the effects of measurement error are quite severe, with the median effect estimates being relatively similar, but the confidence interval around them in this case being considerably larger. The effect estimate is also considerably different from that from a simple OLS model (i.e., $\beta_{simulation}$ = -0.255, while β_{OLS} = -0.542).

Our point in this section is not a criticism of the WGI measure of rule of law. Rather, it is a concern about the ***use*** of this data (or really any other data likely to contain measurement error)—something we ourselves have also done—without ***explicitly*** taking this uncertainty into account. To our knowledge this has not yet been done in any systematic fashion in any of the myriad studies that have employed the WGI. Nor do we mean to imply that the WGI are inferior to similar perceptions-based measures; they may well be better, though it is hard to tell as competing approaches do not provide estimates of uncertainty. What we are concerned about is the degree to which reliance on error- (and bias-) prone measures like this render cross-national analysis difficult-to-impractical.

In the section that follows we lay propose an alternative approach that relies on objective measures of characteristics of institutions and/or institutional outputs in an efficiency sense. While this approach is not without its limitations – most notably that it necessarily considers separately discrete institutions and institutional tasks—it is our contention that it will ultimately be more fruitful insofar as it helps us to overcome the biases and uncertainties that inhere in the regnant perceptual approaches to measuring governance.

4. Conclusion: bringing objectivity back in?

Where do we go from here? We begin to provide an answer by briefly reviewing the experience of one of the more successful subjective measures in the history of the social and medical sciences: self-reported health status. A quarter of a century has passed since Jana Mossey and Evelyn Shapiro first realized that self-reported health status provided a better predictor of seven year survival rates among the Canadian elderly than either data drawn from medical records or self-reports of specific conditions (Mossey and Shapiro 1982) and much has been learned in that time. Subjective assessments of global health status are by now known to: (i) be highly correlated with other indicators (or correlates) of health and well-being; (ii) add enormous explanatory power to multivariate models of mortality (Idler and Benyamini 1997); and (iii) suffer from conceptual limitations that sharply circumscribe their practical utility (Krause and Jay 1994).

Medical professionals know that individuals who label their own health "poor" are likely to die sooner than people who consider themselves fit (Idler and Benyamini 1997), but they don't know why they do so and are therefore unable to act upon their knowledge. Some suspect that self-reports are more inclusive than objective data, tap undiagnosed diseases, or capture the effects of co-morbidity. Others think they capture trajectories rather than levels. Some point to underlying correlations with family history, socio-economic status, or behavioral characteristics. And still others to self-fulfilling prophecies. But unless and until an actual causal mechanism is adduced medical professionals will continue to act not upon subjective but upon objective data when choosing treatment.

Perceptions-based measures of governance have much in common with subjective assessments of health status. They are, after all, highly correlated with each other and with the presumed covariates of good governance (e.g., GDP per capita, school attainment). They, too, add explanatory power to multivariate models. And they also suffer conceptual shortcomings that undercut their practical utility. Take, for example, the rule of law. A positive RL coefficient in a cross-national growth regression is difficult to interpret let alone act upon. It could reflect the influence of one or more of the actual inputs to RL (e.g., crime, courts, contracts), which are themselves endogenous and are therefore not readily altered by public policy. Or it could betray the impact of the direct effects of the social and economic conditions that influence those inputs (i.e., per capita income, social capital, etc.), which are arguably even less susceptible to policy manipulation. It could be a product of pure measurement error of the sort admitted (but perhaps underestimated) by the indicator's authors. Or it could be the product of herd behavior among investors who purchase their information from the indicator's underlying sources. Unless and until we have clear answers to these questions, however, the WGI will be of limited practical utility.

The problem is aggravated, we believe, by the implicitly contradictory targets embedded in the indicators and subindicators that go into the WGI. After all, the sources employed by Kaufmann and his colleagues reward governments for policies that are almost certainly in tension with each other. What, then, should developing country policymakers who hope to improve their WGI scores *do*? Outlaw the expropriation of assets in order to maximize RL or adopt laws like RICO and use them to expropriate corrupt businesspeople and politicians so as to gain better scores on the control of corruption (CC) indicator? Extract and invest resources in schools and roads so as to bolster "government effectiveness" (GE) or cut taxes in order to ensure "regulatory quality" (RQ)? Clamp down on protest so as to ensure "political stability" (PS) or give the protesters and the press a free pass in an effort to maximize "voice and accountability" (VA)? Almost every potential solution aggravates another problem, and the WGIs therefore punish poor countries for their very poverty. If they could solve their social and economic problems, after all, they wouldn't be worried about their governance scores in the first place.

What, then, is to be done? The medical professionals who collect and analyze self-reported health data ultimately turn to objective alternatives when diagnoses and treatments are necessary. Unfortunately, however, objective governance data are not readily available. The traditional proxies—e.g., tax ratios and

the like—have been largely discredited for one reason or another. And nobody has seen fit to invest in the creation of alternatives. We think such an effort is long past due, however, and would yield important payoffs not only for social scientists and policymakers who are skeptical of perceptions-based measures but also, ironically, for their defenders, for the existing subjective indicators arguably cry out for the legitimating aura of an objective benchmark.

What would objective governance indicators look like? First, they would of necessity be issue-specific. While subjective health data are "global" in nature, the objective data that guide diagnosis and treatment are necessarily circumscribed to particular health problems. Similar specification would be required for governance data, especially in light of the fact that governance is known to vary across issue areas and sectors within countries as well as between countries more generally (see, e.g., Johnson 1982). Second, they would focus on issues with measurable outputs: education, health care provision, and postal delivery come immediately to mind (see Putnam 1992). Third, they would incorporate data on *inputs* as well as outputs. One problematic feature of the WGI is their almost complete inattention to the problem of opportunity cost. A government that bankrupted itself in a successful war against corruption would presumably see its corruption control score improve—at least in the short run. But it's not clear that it would have made a wise choice in doing so. And, finally, it would differentiate the efficacy of governance—that is, the enforcement of the rules of the game—from the quality of policymaking—that is, the creation of the rules of the game—as well as from the socio-cultural context that underpins both the rules and their enforcement. It would thus demand independent data on inputs and outputs as well as policies and contextual factors that might affect both.

We can do no more than provide a brutally abbreviated example at present. Neither available data nor the space we have left would tolerate more. But we'll try to illustrate the *sort* of approach we think desirable by examining the issue of labor law enforcement in Latin America—which at least in theory meets the aforementioned criteria. It is not only narrow in scope, at least when compared to "global" issues like the rule of law, but informality provides a straightforward indicator of the output. Successful enforcement should issue in a relatively low level of informality *ceteris paribus*. The number of enforcement personnel constitutes the input. And the ratio of enforcement personnel to the active labor force constitutes the standard measure of the intensity of the enforcement effort (ILO 2004, p. 13). Control variables are available to capture the effects of policy differences and social structure. And residuals from a regression of informality on the aforementioned indicators should therefore provide an admittedly crude proxy for the *quality* of the enforcement effort.

We implement the strategy by regressing the informal percentage of the labor force (Gasparini and Tornarolli 2007, Table 3.3) on GDP per capita (World Bank 2006)—which serves as a simple proxy for the overall level of social and economic development—and an index of job security that is positively, if inconclusively, related to informality (Heckman and Pagés-Serra 2000). Figure 4 plots the cross-national variance in informality unexplained by the combination of GDP per capita and job security regulation by the ratio of labor inspectors to laborers in 17 Latin American countries in the early 2000s (Schrank and Piore 2007, Table 5; and updated data where available).[14]

PLEASE INSERT TABLE 4 ABOUT HERE

The data in the figure suggest several interesting conclusions about governance. First, they fly directly in the face of the widespread notion that "limited government," in Kaufmann et al.'s formulation, is necessarily the best government. While the cross-sectional data permit no causal claims, they are at least

[14] Guatemala is dropped due to missing job security index data. We present the results this way for clarity of exposition. Alternative specifications (e.g., entering all three predictors simultaneously; using different indicators of informality, dropping the job security index and including Guatemala, etc.) yield parallel results.

consistent with the idea that labor market regulations inhibits, rather than promotes, informality in Latin America. Second, they suggest that quality matters too. The three largest negative residuals—Chile, Costa Rica, and the Dominican Republic—have all taken efforts to improve the quality of enforcement in recent years (see Schrank and Piore 2007). By way of contrast, two of largest positive residuals, Brazil, and in particular Mexico, have moved in the opposite direction. And, third they suggest that objective governance indicators could, with a good deal of additional effort, provide a valuable complement or even alternative to the existing perceptions-based measures. One would not only need longitudinal as well as cross-sectional data from a larger number of countries but would also have to deal with questions of endogeneity and selection. But these issues arise with subjective data as well—and there they're compounded by issues of conceptual ambiguity, systematic and random measurement error, and causal indeterminacy.

A thoroughgoing collective effort to aggregate and disseminate objective data would not only help researchers combat such problems but would also provide an alternative if such a campaign were to fail. Data on inputs and outputs are available across a wide array of issue areas in a large number of developing countries. It's high time that intergovernmental organizations began to aggregate them and make them available to policymakers, scholars, and activists around the world.

References

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Quanlitative and Quantitative Research." *American Political Science Review.* 95 (3): 529-46.

Ades, Alberto, and Rafael DiTella. 1995. "National Champions and Corruption: Some Unpleasant Interventionist Arithmetic." *Economic Journal.* 107 (443): 1023-42.

Apodaca, Claire. 2004. "The Rule of Law and Human Rights." *Judicature.* 87 (6): 292-9.

Arndt, Christiane, and Charles Oman. 2006. *Uses and Abuses of Governance Indicators.* Paris: OECD.

Arnold, Wayne. 1999. "Tough Guy at the Loan Office; Backed by I.M.F., a Reformer Tackles Indonesia's Barons." *New York Times.* July 28..

Benson, Michelle, and Jacek Kugler. 1998. "Power Parity, Democracy, and the Severity of Internal Violence." *Journal of Conflict Resolution.* 42 (2): 196-209.

Bergman, Marcelo. 2006. "Crime and Citizen Security in Latin America: The Challenges for New Scholarship." *Latin American Research Review.* 41 (2): 213-27.

Bhagwati, Jagdish. 2007. "Comments." *Financial Times.* FT Economists' Forum. November 7. http://blogs.ft.com/wolfforum/2007/11/why-plutocracy-.html#comments

Blake, Charles, and Christopher Martin. 2006. "The Dynamics of Political Corruption: Re-examining the Influence of Democracy." *Democratization.* 13 (1): 1-14.

Blumenson, Eric, and Eva Nilson. 1998. "Policing for Profit: The Drug War's Hidden Economic Agenda." *University of Chicago Law Review.* 65 (1): 35-114.

Borrrmann, Axel, Matthias Busse, and Silke Neuhaus. 2006." Institutional Quality and the Gains from Trade." 59 (3): 345-58.

Byerly, Henry. 1972. "Realist Foundations of Measurement." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association.* 375-82.

Cameron, Maxwell, Ana-Maria Blanaru, and Lesley Burns. 2006. "Constitutional Frameworks and the Rule of Law: Between Delegative Democracy and Elective Dictatorship." Paper presented to the American Political Science Association. Philadelphia. September.

Carnap, Rudolf. 1962. *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Daniels, Ronald, and Michael Trebilcock. 2004. "The Political Economy of Rule of Law Reform in Developing Countries." *Michigan Journal of International Law.* 26: 99-108.

Dumont, Richard, and Wilson, William. 1967 "Aspects of Concept Formation, Theory Contsruction, and Explication in Sociology." *American Sociological Review.* 32 (6): 985-95.

Economist. 1981. "Chile: capitalists gaffed." December 12.

Evans, Peter. 1997. "The Eclipse of the State? Reflections on Stateness in an Era of Globalization." *World Politics.* 50 (October): 62-87.
Fish, M. Steven. 2003. "The Hazards of Half Measures: Perestroika and the Failure of Post-Soviet Democratization." *Demokratizatsiya.* 13 (2): 241-53.
Gasparini, Leonardo, and Leopoldo Tornarolli. 2007. "Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata." CEDLAS Working Paper No. 0.
Gerring, John, and Strom Thacker. 2005. "Do Neoliberal Policies Deter Corruption?" *International Organization.* 59 (Winter): 233-54.
Gervasoni, Carlos.2006. "Data Set Review: The World Bank's Governance Indicators (1996-2004)." *APSA-CP Newsletter.* 17 (1): 17-20.
Goldsmith, Arthur. 1995. "Democracy, Property Rights and Economic Growth." *Journal of Development Studies.* 32 (2): 157-174.
Goldsmith, Arthur. 1999. "Slapping the Grasping Hand: Correlates of Political Corruption in Emerging Markets." *American Journal of Economics and Sociology.* 58 (4): 865-883.
Heckman, James, and Carmen Pagés-Serra. 2000. "The Cost of Job Security Regulation: Evidence from Latin American Labor Markets." *Economia.* Fall.
Hempel, Carl. 1956. "Logical Appraisal of Operationism." *Scientific Monthly.*
Idler, Ellen, and Yale Benyamini. 1997. "Self-Rated Health and Mortality: A Review of Twenty-Seven Community Studies." *Journal of Health and Social Behavior.* 38 (1): 21-37.
International Labour Organization. 2004. *Statistical Activities of the International Labour Organization.* Geneva: ILO.
Jacobs, James, and Lauryn Gouldin. 1999. "Cosa Nostra: The Final Chapter?" *Crime and Justice.* 129-99.
Johnson, Chalmers. 1982. *MITI and the Japanese Miracle: The Rise of Industrial Policy.* Berkeley: University of California Press.
Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999a. "Aggregating Governance Indicators." World Bank Institute.
Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobatón. 1999b. "Governance Matters." World Bank Institute.
Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005a. "Measuring Governance Using Cross-Country Perceptions Data." World Bank Institute.
Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005a. "Governance Matters IV." World Bank Institute.
Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2006. "Governance Matters V." World Bank Institute.
Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2007a. "Growth and Governance: A Reply." *Journal of Politics.* 69 (2): 555-62.
Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2007b. "Governance Matters VI." World Bank Institute.
Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2007c. "The Worldwide Governance Indicators Project: Answering the Critics." World Bank Institute. February.
Kaufmann, Daniel, and Aart Kraay. 2002. "Growth without Governance." *Economica.* 3 (1): 169-229.
Kling, Merle. 1959. "Taxes on the "External" Sector: An Index of Political Behavior in Latin America?" *Midwest Journal of Political Science.* 3 (2): 127-50.
Knack, Stephen. 2007. "Measuring Corruption: A Critique of Indicators in Eastern Europe and Central Asia." *Journal of Public Policy.* 27 (3): 245-91.
Krasner, Stephen. 1985. *Structural Conflict: The Third World Against Global Liberalism.* Berkeley: University of California Press.
Krause, Neal, and Nina Jay. 1994. "What Do Global Self-Rated Health Items Measure?" *Medical Care.* 32 (9): 930-42.

Kurtz, Marcus, and Andrew Schrank. 2007a. "Growth and Governance: Models, Measures, and Mechanisms." *Journal of Politics.* 69 (2): 538-54.
Kurtz, Marcus, and Andrew Schrank. 2007b. "Growth and Governance: A Defense." *Journal of Politics.* 69 (2): 563-69.
Laurence, Charles. 2004. "Mafia's 'last godfather' goes on trial for murder and racketeering." *Sunday Telegraph.* May 16.
Lee, Cheol-Sung. 2007. "Labor Unions and Good Governance: A Cross-National, Comparative Analysis." *American Sociological Review.* 72 (4): 585-609.
Levy, Leonard. 1996. *A License to Steal: The Forfeiture of Property.* Chapel Hill: UNC Press.
Lewis, J.B. and D.A. Linzer. 2005. "Estimating regression models in which the dependent variable is based on estimates." *Political Analysis.* 345-64.
Mauro, Paolo. 1995. "Corruption and Growth." *Quarterly Journal of Economics.* 100 (3): 681-72.
Migdal, Joel. 1988. *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World.* Princeton: Princeton University Press.
Mossey, Jana, and Evelyn Shapiro. 1982. "Self-Rated Health: A Predictor of Mortality among the Elderly." *American Journal of Public Health.* 72 (8): 800-808.
Newseek. 1989. "The RICO Weapon: Too Harsh for Wall Street?." January 2.
Organski, A.F.K., and Jacek Kugler. 1980. *The War Ledger.* Chicago: University of Chicago Press.
Portes, Alejandro, and Brian Roberts. 2005. "The Free Market City: Latin American Urbanization in the Years of the Neoliberal Experiment." *Studies in Comparative International Development.* 40 (1): 43-82.
Przeworski, Adam, and Joseph Sprague. 1971. **"**Concepts in Search of Explicit Formulation: A Study in Measurement **."** *American Political Science Review.* 15 (2): 183-218.
Reynolds, Lloyd. 1983. "The Spread of Economic Growth to the Third World: 1850-1980." *Journal of Economic Literature.* 21(3): 941-80.
Rigobon, Roberto, and Dani Rodrik. 2005. "Rule of Law, Democracy, Openness and Income: Estimating the Interrelationships." *Economics of Transition.* 13 (3): 533-564.
Ringer, Thom. 2007. "Development, Reform, and the Rule of Law: Some Prescriptions for a Common Understanding of the 'Rule of Law' and its Place in Development Theory and Practice." *Yale Human Rights and Development Law Journal.* 10: 78-208.
Rodrik, Dani. 2007. "What does the "rule of law" mean?" Weblog entry. November 13. http://rodrik.typepad.com/dani_rodriks_weblog/2007/11/what-does-the-r.html
Salmon, Wesley. 1999. "The Spirit of Logical Empiricism: Carl G. Hempel's Role in Twentieth-Century Philosophy of Science." *Philosophy of Science.* 66 (3): 33-50.
Sampson, Robert, Stephen Raudensbuh , and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science.*277: 918-24.
Sandholtz, Wayne, and Mark Gray. "International Integration and National Corruption." *International Organization.* 57 (Autumn): 761-800.
Schrank , Andrew. "Asian Industrialization in Latin American Perspective: The Limits to Institutional Analysis." *Latin American Politics and Society.* 49 (2): 183-200.
Schrank, Andrew, and Michael Piore. 2007. "Norms, regulations and labor standards in Central America." Mexico City: CEPAL.
Snider, Lewis. 1988. "Comparing the Strength of Nations: The Arab Gulf States and Political Change." *Comparative Politics.* 20 (4): 461-84.
Stephenson, Matthew. 2000. "A Trojan Horse Behind Chinese Walls?: Problems and Prospects of US-Sponsored 'Rule of Law' Reforms in the People's Republic of China." Harvard. CID Working Paper No. 47. Law and Development Paper No. 3.
Thomas, Melissa. 2007. "What do the Worldwide Governance Indicators measure?" Johns Hopkins SAIS. August.
Treier, Shawn, and Simon Jackman. 2007. "Democracy as a Latent Variable."
World Bank. 2006. *World Development Indicators.*

Yatchew, Adonis, and Zvi Griliches. 2005. "Specification Error in Probit Models." *Review of Economics and Statistics.* 134-9.

Table 1. Components of the WGI Rule of Law Indicator

Rule of law—measuring the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence

	Source	Type	Respondents	Questions
Representative	Global Insight Global Risk Service	Commercial information provider	Staff	Costs of crime Kidnapping Contract enforcement (public and private)
	Economist Intelligence Unit	Commercial information provider	Correspondents	Violent crime Organized Crime Judicial speed/fairness Contract enforcement Expropriation
	World Economic Forum Global Competitiveness Survey	Nongovernmental organization (with business representation)	Firms (survey)	Common crime (cost to business) Organized crime (cost to business) Money laundering Police Quality Judicial independence and reliability Intellectual property rights Financial protection Tax evasion Illegal political donations
	Gallup World Poll	Commercial polling firm	Households (survey)	Confidence in police Confidence in courts Crime victimization
	Heritage Foundation Index of Economic Freedom	Nongovernmental organization (conservative)	Staff	Property Rights
	Cingranelli & Richards Human Rights Database	Academics	Expert codings of US Dept. of State and Amnesty International reports	Judicial independence
	Merchant International Group Gray Area Dynamics	Commercial information provider	Staff	Organized crime Legal safeguards
	Political Risk Service International Country Risk Guide	Commercial information provider	Staff	Law and order
	Business Environment Risk Intelligence Quantitative Risk Measure in Foreign Lending	Commercial information provider	Staff	Financial fraud Money laundering Organized crime
	US Department of State Trafficking in Persons Report	USDOS	Expert assessments	Human trafficking
	Global Insight Business Conditions and Risk Guide	Commercial information provider	Staff	Judicial independence Crime
Non-representative	African Development Bank Country Policy and Institutional Assessments	African Development Bank	Country economists	Property rights
	Afrobarometer	University/NGO collaboration	Household survey	Police response time
	Asian Development Bank Country Policy and Institutional Assessments	Asian Development Bank	Country economists	Rule of law
	Business Environment and Enterprise Performance Surveys	World Bank & European Bank for Reconstruction and Development	Survey (firms)	Judicial quality Property rights Crime and business
	Business Environment Risk Intelligence	Commercial information provider	Panel of experts	Contract enforcement
	Bertelsmann Transformation Index	Nongovernmental organization	Staff	Rule of law Private property
	Freedom House Countries at the Crossroads	Nongovernmental organization	Staff and consultants	Rule of law
	World Bank Country Policy and Institutional Assessments	World Bank	Country Economists	Property rights
	Freedom House	Nongovernmental organization	Staff and consultants	Rule of law
	Global Integrity Index	Nongovernmental organization	Local country experts and peer review	Executive accountability Judicial accountability Rule of law Law enforcement
	IFAD Rural Sector Performance Assessments	IFAD	Country economists	Access to land Access to agricultural water
	Latinobarometer	Nongovernmental organization	Survey (household)	Trust in judiciary Trust in courts Crime victimization
	Institute for Management Development's World Competitiveness Yearbook	Education organization	Survey (business)	Judicial quality Intellectual property Crime

Source: Adapted from Kaufmann et al. 2007b, esp. Table B5. Representative sources are available for most countries and therefore weigh more heavily in the final indicators.

Table 2: Do businesspeople and their neighbors perceive government in the same way?

Question	Responses	Businessperson
Trust in the judiciary	1 = much trust; 2= some trust; 3= little trust; 4 = no trust	Odds ratio = 1.15 ($p < .010$)
Trust in the police		Odds ratio = 1.23 ($p < .001$)

Self-identified businesspeople are coded 1; others are coded 0; country dummies are suppressed; and odds ratios for businesspeople are presented next to their parenthesized p values. Non responses and "don't know" are dropped. The data are from Latinobarometer (2005); the more recent data available to KKM are not publicly available.

Table 3. Estimated Correlations among Repeated Observations of the Rule of Law, 1996
(Correlation over 95 percent confidence interval)

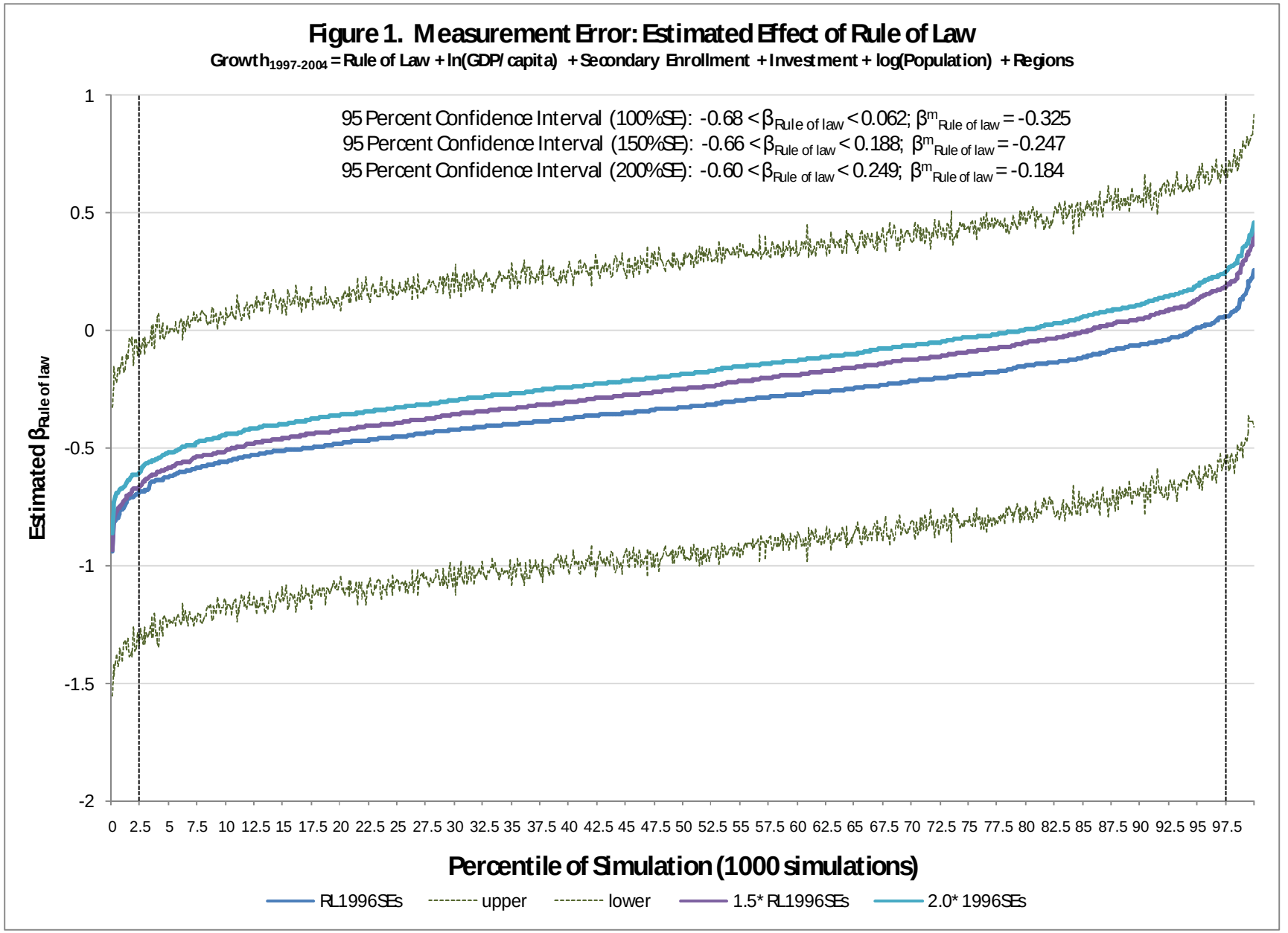
All countries

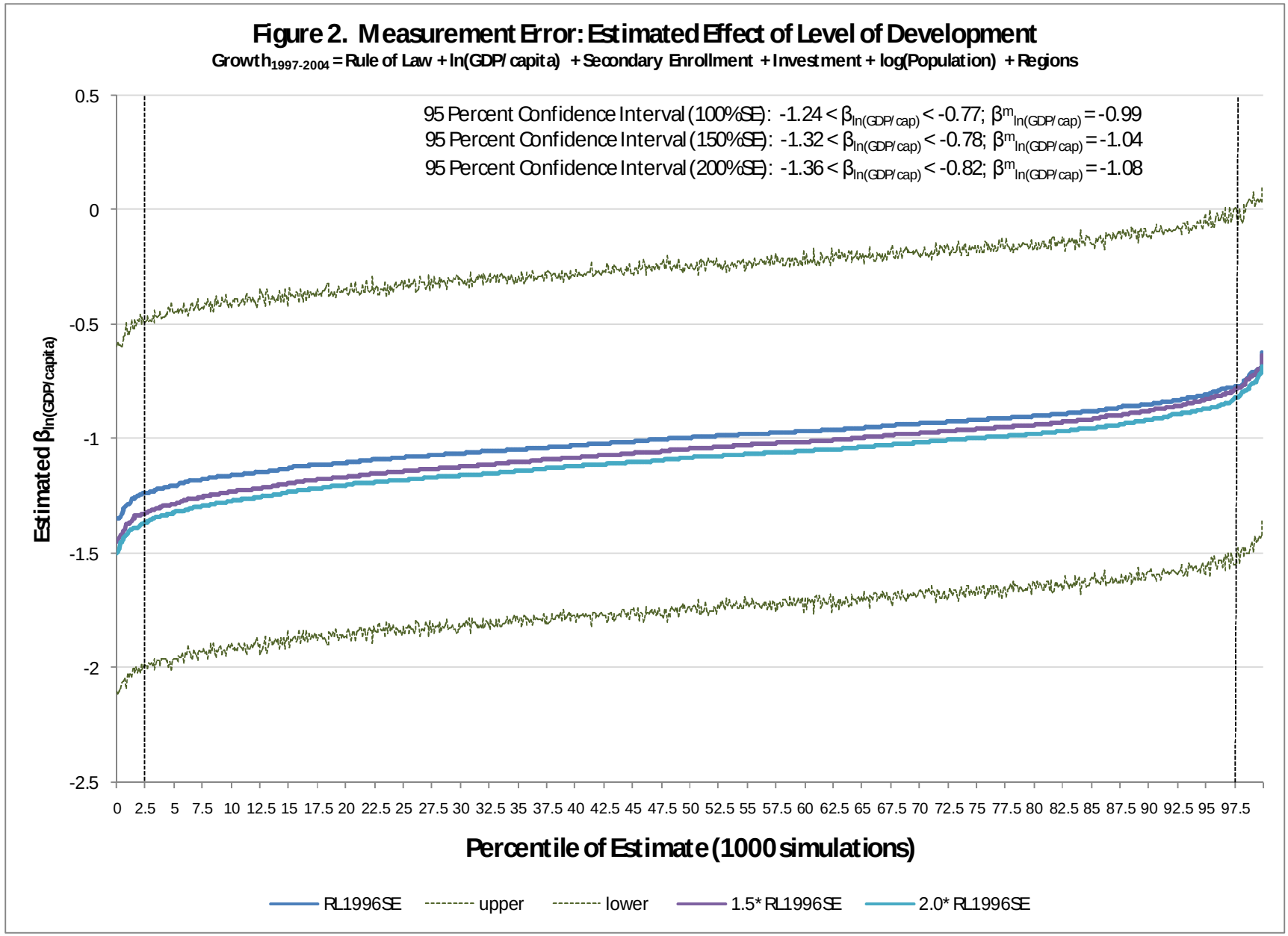
	Rule of Law, 1996 KKM SEs	Rule of Law, 1996 1.5*KKM SEs	Rule of Law, 1996 2.0*KKM SEs
Correlation	0.909 (median)	0.818 (median)	0.716 (median)
95% Confidence Interval	[0.876 – 0.934]	[0.752 – 0.867]	[0.607 –0.788]
Simulations	1000	1000	1000

For the countries below the median level of GDP/capita

	Rule of Law, 1996 KKM SEs	Rule of Law, 1996 1.5*KKM SEs	Rule of Law, 1996 2.0*KKM SEs
Correlation	0.724 (median)	0.538 (median)	0.398 (median)
95% Confidence Interval	[0.581 – 0.821]	[0.343 – 0.701]	[0.148 –0.600]
Simulations	1000	1000	1000

Notes: Monte Carlo simulation of rule of law data using KKM estimates and standard errors to draw repeated observations, assuming a normal distribution of mean zero and standard deviation as given by the KKM standard errors and multiples thereof.





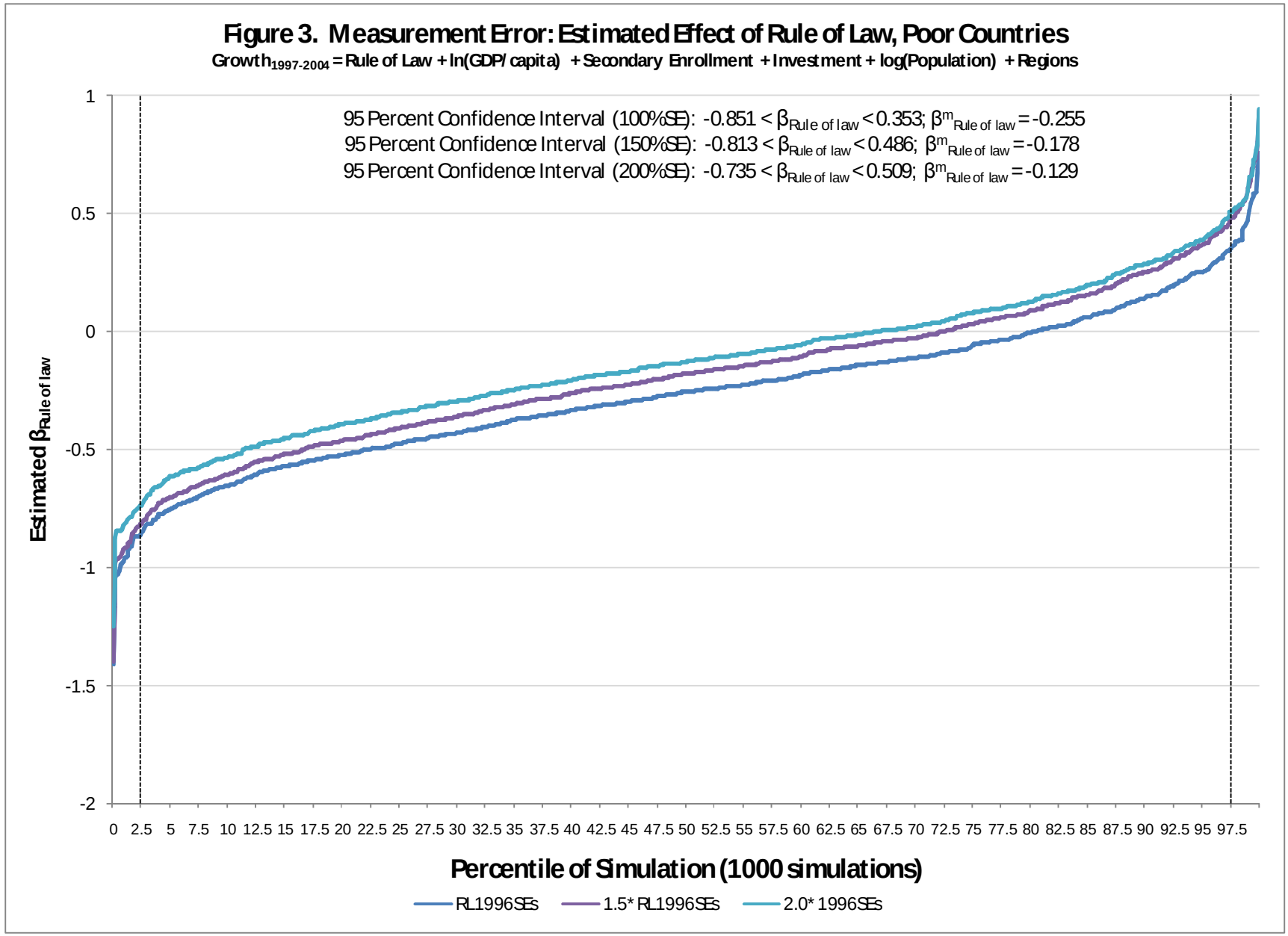


Fig. 4: Labor law enforcement in Latin America: quantity and quality