WPS4669

POLICY RESEARCH WORKING PAPER 4669

The Worldwide Governance Indicators and Tautology:

Causally Related Separable Concepts, Indicators of a Common Cause, or Both?

Laura Langbein
Stephen Knack

Abstract

Aggregate indexes of the quality of governance, covering large samples of countries, are widely used in research and in aid policy. Few studies examine the validity of these indexes, however. This paper partially fills this gap by examining empirically the dimensionality of the Worldwide Governance Indicators. The six indexes purportedly measure distinct concepts of control of corruption, rule of law, government effectiveness, regulatory quality, political stability, and voice and accountability. Using standard statistical techniques for testing measurement validity, the analysis concludes that the six indexes do not discriminate usefully among different aspects of governance. Rather, each of the indexes merely reflects perceptions of the quality of governance more broadly. An implication of the findings is that the Worldwide Governance Indicator indexes are frequently misused in research and policy applications, where it is commonly assumed that the indexes provide distinct measures of different aspects of the quality of governance. A further implication is that Transparency International's even more widely-known aggregate index similarly reflects perceptions not only of corruption, as intended, but of the quality of governance more broadly.

The Worldwide Governance Indicators and Tautology: Causally Related Separable
Concepts, Indicators of a Common Cause, or Both?

Laura Langbein
Dept. of Public Administration and Policy
American University School of Public Affairs
4400 Massachusetts Av., NW
Washington, DC 20016
202-885-6233 (ph) 202-885-2347 (fax)
langbei@american.edu

Stephen Knack
Lead Economist
Development Research/Public Sector Governance
The World Bank
MSN MC3-313
1818 H St. NW
Washington DC 20433
202-458-9712 (ph) 202-522-1154 (fax)
sknack@worldbank.org

The Worldwide Governance Indicators and Tautology:
Causally Related or Common Cause?

Introduction

In recent years aggregate indexes of the quality of governance, covering large samples of countries, have become enormously popular in comparative political analysis. However, few studies have examined the validity or reliability of these indexes (Thomas, 2007). This study attempts to partially fill this gap. We find that these indexes, purportedly measuring distinct concepts such as control of corruption and rule of law, all appear to be measuring essentially the same broad concept, rather than successfully distinguishing among various aspects of the quality of governance.

Beginning in 1995, Transparency International has annually produced a corruption index covering a large sample of countries (Lambsdorff, 1998). The index does not represent TI's own assessments, but is simply aggregated from numerous other sources, including expert opinions and surveys of firms and households.

Several researchers at the World Bank adopted this basic approach of the TI index, but attempted to improve on it in several respects (Kaufmann, Kraay and Zoido-Lobaton, 1999) in their "Worldwide Governance Indicators" (WGI) project. Most notably, they have exploited the data sources more fully to produce six measurement indexes: in addition to Control of Corruption (CC), these include Voice and Accountability (VA), Rule of Law (RL), Government Effectiveness (GE), Political Stability (PS) and Regulatory Quality (RQ).

There are various ways of reducing the vast content of the numerous available data on quality of governance into a smaller number of aggregate indexes. The six WGI indexes were selected to measure separate, but related, concepts regarding the quality of governance. The content of each of them was determined purely on conceptual grounds, and not by empirical means such as factor analysis.

Yet the task of establishing measurement reliability and validity is not logically any different from that of testing theories about cause and effect. Theories of cause and effect should not only be logically persuasive; they should also be empirically convincing. It is conventional to evaluate the defensibility of a logically persuasive

empirical claim regarding causality using criteria of internal and statistical validity. For example, we seek parameter estimates that are as unbiased and efficient as reasonably possible. If empirical evidence systematically fails to support a seemingly logical causal claim, we are likely to rethink our original logic.

The same is true about measurement. A claim about how to measure an abstract concept may appear logically or conceptually convincing, but claims about measurement should (and can) also be tested empirically, using the same criteria of internal and statistical validity that are conventionally applied to testing claims about causality. In the context of measurement, we want the indicators of an abstract concept to systematically and reliably relate to that concept (and not other, different, concepts). Respectively, this means that we seek indicators that measure the hypothesized abstract concept with minimal systematic (non-random) and random error.

There is little if any evidence on the concept validity of the six WGI indexes. Concept validity requires that an indicator of abstract concept A should be systematically related to concept A and not related to concept B. Similarly, the indicator of abstract concept B should be much more closely related to concept B than to concept A.

The six WGI indexes are often treated as measuring six distinct concepts. For example, the 18 eligibility criteria considered by the U.S. Millennium Challenge Corporation (MCC) in allocating aid include not simply an average of all of the WGI indexes, but five of them separately, listed by MCC under two separate conceptual categories ("ruling justly" and "economic freedom"). A large and growing number of research papers employ one of the six indexes to test fairly specific hypotheses, in most cases without acknowledging the possibility that the variable is really reflecting broader concepts related to the quality of governance.

For example, some analyses include one WGI index (or component measures) on the left-hand-side and one or more other WGI indexes (or component measures) on the right-hand-side. Several studies (e.g. Tavits, 2007; You and Khagram, 2005; Sandholtz and Koetzle, 2000) use CC or TI's corruption index as a dependent variable and the Freedom House political rights index as an independent variable. The latter's criterion includes "is the government free from pervasive corruption?" In recent years the scoring even takes into account the TI corruption index itself. In Damania, Fredriksson and Mani

(2004), political instability impairs rule of law, in turn stimulating corruption. These concepts are operationalized by the WGI indexes PS, RL and CC. In Brewer, Choi and Walker (2007), CC and VA both affect GE.[1]

Other researchers treat the WGI indicators as distinct outcomes, interpreting separate regressions as distinct and additional pieces of evidence. For example, Alence (2004) find that democratic contestation and executive restraints affect RQ, GE and CC. Others use the governance indicators to investigate simultaneous relations between income or GDP growth or other measures of government performance and one or more of the WGI indexes (Kaufman and Kraay, 2002; Ritzen, Easterly and Woolcock, 2000). This may be particularly problematic since the WGI indicators are largely perceptual, and a strong economy can elicit responses affirming good governance (Kurtz and Schrank, 2007). The indexes are also used by aid practitioners to assess the relative strengths and weaknesses of a particular country among these six broad dimensions of governance. The validity of this sort of diagnosis rests on the ability of the WGI indexes to discriminate effectively among these six concepts, and to be different from other measures of government performance.

This study examines empirically the dimensionality of the WGI indexes. We test whether the six governance indicators measure a broad underlying concept of "effective governance," or whether they are separate, causally related concepts. We find that the indicators are consistent with both; that is, they represent a single underlying concept AND they are causally related, separate indexes. In effect, they appear to say the same thing, with different words. Based on this evidence of tautology we conclude that the six indexes do not discriminate usefully among different aspects of governance. Rather, each of the indexes – whatever its label – merely reflects perceptions of the quality of governance more broadly. An implication is that they may have limited use as guides for policymakers, and for academic studies of the causes and consequences of "good governance" as well.

---

[1] After reporting their empirical results, however, they suggest that all of the indexes may be "tapping the same topics and concepts" and that the various WGI indexes may have highly correlated measurement errors because "experts are likely to share perceptions and read the same reports to assess complex concepts."

Theory: The indicators

The WGI indexes are designed to signify the relative absence or presence of some very closely related phenomena. In each case, they appear either to be causally related, or related by definition. They are intended to measure the relative presence of the following properties.

Voice and accountability (VA). Corruption (CC) is usually hidden, which is part of its usual definition. Hidden transactions imply weaknesses in transparency and accountability. The VA index is intended to measure the ability of citizens to hold politicians accountable, including freedom of press, association, and media. Thus, conceptually, VA and CC are either causally related or are related by definition.

Government effectiveness (GE). Effective governments make transfers, but they are not hidden (VA). Also, effective governments use public resources, often for public gain, so that the spending is not a deadweight loss (rule quality, or RQ). Effective governments charge for services that citizens want, implying again no or minimal deadweight loss. The GE index is intended to measure the ability of governments to deliver these public services, the quality of the civil service (the "agents" of delivery), and the independence of the bureaucracy from political influence, including the credibility of bureaucratic commitment to its policies ("unbribe-ability," implicating CC). There is thus a causal or definitional overlap among the three concepts GE, RQ, and CC.

Rule of law (RL). Corrupt deals are enforced in a black market, where contracts are enforced not by public law but by private bandits. Rule of law implies an open, "white", transparent market, where contracts are enforced by a "rule" that is publicly known to parties outside the contract and applied equitably no matter who the enforcer or the contract-parties are. The variable is intended to measure the probability that contracts and laws or rules are enforced collectively, and accountably, rather than privately. Thus, RL and VA are connected, by cause or by definition. Corrupt activities are typically illegal, indicating rule of law weaknesses. Thus, RL and CC are related by definition, if not by cause.

Political stability (PS). Corruption (CC) is a pure social waste. Governments, political parties, and political officials with a long time horizon (PS) know that economic growth is the friend of longevity, and they will not support highly ineffective government

(GE) and an excess of rules, and prefer the rule of law to the rule of bandits (RL). When government transitions are decided by well-defined and long-lived rules, rather than by violent overthrow or perennial coups, government officials are more likely to have a longer time horizon, and to seek investment for growth (and political survival) rather than corrupt transfers. The PS index is intended to measure the expected orderliness of political transitions according to established rules. Thus, PS is related to CC, RL and GE either causally or by definition.

Control of corruption (CC). Controlling corruption implies a reduction in the use of public resources for private gain. Corruption is a costly, hidden (absence of VA) and usually illegal (absence of RL) transfer of revenues. Government officials often collect bribes as an ex-officio tax, fee or "gift" in exchange for a license or service (e.g. utility connection), or for exemptions to rules or taxes (implicating RQ and GE).

Regulatory quality (RQ). When governments establish numerous barriers to conducting business, it creates opportunities for public officials to collect bribes before delivering a service (CC). By definition, corrupt governments set up entry barriers so that public officials can act as gate-keepers, and collect (hidden) bribes and pocket the transfer before opening the gate to the briber-client (absence of VA). (There can also be too few rules: higher-quality regulation implies there are not excessive rules, and that rules are efficiency enhancing.) The RQ index is intended to measure the extent to which the formal (and informal) regulations that govern the relation between the public and the private sector (RL) foster growth rather than costly transfers (from the private client to the public regulator, or the other way). Thus RQ implicates (by definition or cause) CC, RL, RQ and VA.

In operationalizing these concepts empirically, there are additional obstacles preventing the WGI indexes from successfully representing six distinct concepts. First, in collapsing the large number of "governance" indicators they collected into a small number of groups, Kaufmann and his colleagues may have selected the wrong six concepts, or they may have chosen the wrong number of them. Arguably, they should have supplemented their intuition with some exploratory empirical work in identifying dimensions, but they do not report having done this analysis. The breadth of some concepts is potentially problematic, with "government effectiveness" the most obvious

6

example of an overly broad, vague and heterogeneous concept that may be impossible to distinguish from other aspects of good governance.

Second, some of the six WGI concepts overlap sufficiently so that assignment of some of the available component indicators to one index or another is essentially arbitrary. For example, component indicators measuring constraints on the executive or human rights could be assigned either to RL or to VA. Indicators on perceptions of violent or non-politically-motivated crime could equally be assigned to RL or to PS, which is defined not only in terms of political stability but also absence of violence. The Freedom House "political rights" and "civil liberties" indicators are both assigned to VA, but they contain criteria relevant to all five of the other WGI indexes. "Political rights" includes an assessment of corruption (CC). "Civil liberties" includes an assessment of rule of law (RL) and of the population's freedom from "physical harm, forced removal, or other acts of violence or terror due to civil conflict or war" (PS). Its criteria even cover competition policy, free markets in land, price controls and production quotas, and compensation in eminent domain proceedings (RQ). The WGI authors assign a World Bank indicator called "transparency, accountability and corruption" to CC, but the indicator's criteria is equally relevant to VA. Indicators of "red tape" and time spent by business managers in dealing with government officials are assigned to GE, but could as easily have been assigned to RQ. A survey measure on terrorism and crime is assigned to PS, but could as easily have been included in RL.

Third, even if six conceptually meaningful and distinct concepts can be defined, and are followed in principle by the data sources, limitations of the information available to the data sources may prevent them from discriminating successfully among the concepts in operational measures. For example, risk rating services or other sources of subjective governance ratings may sometimes infer in the absence of detailed knowledge that "governance" in all its dimensions must not be too bad if the economy is growing (Kurtz and Shrank, 2007). These sorts of inferences could be particularly common in rating small developing countries for which real information on governance may be very limited.

Finally, the WGI authors occasionally assign component indicators to their six aggregate indexes in erroneous or inconsistent ways. Although the Economic

Intelligence Unit of the Word Bank includes its "orderly transfers" indicator as part of its own "political stability" sub-index, the WGI authors assign it to VA instead of to PS. Similarly, indicators of "institutional permanence" from Global Insight and "institutional stability" from Bertelsman are assigned to VA instead of to PS. But an indicator on "government stability" from the International Country Risk Guide is assigned neither to PS nor to VA, but to GE. Some indicators on tax policy and administration are assigned to GE, but others are assigned to RQ. For several years they assigned the Heritage Foundation's "informal market activity" rating to RL. The extent of informal market activity is often used as a proxy for regulatory problems, including inefficient tax structure and weaknesses in tax administration, so conceptually it could belong in RQ or GE instead of in RL. However, Heritage "measures" its informal market activity indicator simply by re-scaling Transparency International's corruption index. Presumably the WGI authors simply did not read the Heritage methodology paper, or they would not have included it in RL.[2]

　　We ask, and try to answer, the following question: Are these really separate, causally related, concepts, or are they part of the general concept (or definition) of "good government?" Is "failed" government a syndrome that manifests itself in governments that are unaccountable, ineffective, ruled by "bandits", instable, and perceived as "corrupt" in mass and elite surveys? Or, are these separable "causes" of corruption? The answer depends on the how these indicators relate to one another, and to the latent concept of "good government". There are two basic stories of how the six concepts are related: one is causal, and one is a story of measurement. We first examine the support for these two contrasting models; then we consider a mixed model. Figure 1 shows the causal model.

---

[2] In its most recent Index of Economic Freedom, Heritage re-named "informal market activity" to "freedom from corruption," reflecting more accurately the way the rating is produced. The WGI authors no longer use it in their indexes.

Figure 1: The causal model
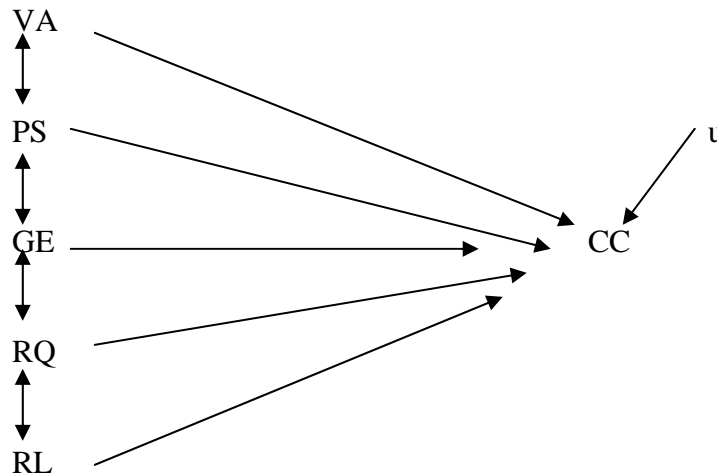
In this causal model, five of the indicators are each a separate concept, but each concept is related to each other concept, and each concept is exogenous to the dependent variable, which is a function of the independent variables and an error term. The error term is independent of each exogenous variable. The conceptual problem with this causal story is that it is not clear that CC should be the dependent variable; GE is a likely candidate for a dependent variable too. This suggests that the error term is unlikely to be independent of each exogenous variable.

Figure 2 shows the measurement model. It requires each measured component to be spuriously related to the common underlying, unmeasured (latent) concept or factor. Each measured component is a linear combination of the same unmeasured factor and an (uncorrelated) error term. The conceptual problem with this story is that it is unlikely that each measured concept has no direct impact on any of the other measured concepts. For example, it is likely that VA, RL and CC are directly (and not spuriously) related to one another.

9

Figure 2: Measurement Model

Good Government

| VA | PS | RQ | RL | GE | CC |

e_{va} e_{ps} e_{rq} e_{rl} e_{ge} e_{cc}
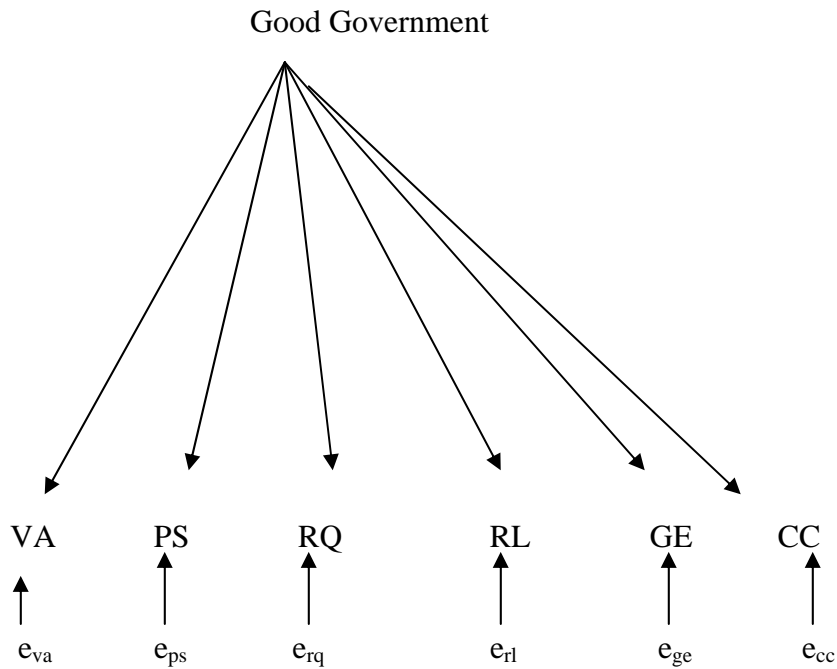
Results:

Preliminaries: The correlations

      First, it is important to report that the 6 WGI variables, based on data for up to 6 years for 216 countries, show vary high bivariate correlations. (See Table 1.) The smallest correlation (r=.64) is between political stability (PS) and voice and accountability (VA), and the largest (r=.91) is between rule of law (RL) and control of corruption (CC). While these correlations seem high, it is important to point out that, in terms of shared variance, there is actually not much overlap between PS and VA, since r^2 = .41. In fact, of the 15 pairs of correlations, only the 6 pairs involving rule quality (RQ), government effectiveness (GE), rule of law (RL) and control of corruption (CC) share more than half their variance.

<u>The Measurement Model</u>

We test the measurement model in three ways. First, if the measurement model is correct, simple exploratory factor analysis should detect one principle factor with a dominant eigenvalue that explains most of the factor space. Each measured variable should correlate highly with that factor, and should not correlate highly with any other factor, and the fit of a single factor model should be better than the fit of a multi-factor model.   Second, if the measurement model is correct, a path analysis of the factor loadings (which correspond to correlations, assuming the underlying model is correct) should predict the following results:

$$r_{ij} = r_{ik} * r_{kj}$$

In other words, if the factor analytic model is correct, the $(6*5)/2 = 15$ observed pairwise correlations among the 6 measured indicators should each equal the product of the two estimated correlations (factor loadings, in this case) between the observed indicator and the underlying single factor. Third, if the measurement model is correct, a confirmatory factor analysis predicting a one-factor solution should produce a good fit to the observed data.

Exploratory (Simple) Factor Analysis:

The results from a principal component factor model show that there is clearly one dominant factor. The eigenvalue from the first factor is 4.58, while the eigenvalue for the second is only .14. (See Table 2.) The factor loadings are also consistent with a one-factor model. All 6 measured variables correlate with the single factor with loadings over .75; the unexplained variance ("uniqueness") is consistently less than 40%. On the other hand, as one might expect from the simple correlation matrix in Table 1, the PS and VA have the lowest factor loadings and the highest unexplained variance, while GE, RQ, RL, and CC have the highest correlation with the dominant factor, and the smallest uniqueness component.

These results do not depend on the method of factor analysis. Extracting factors with maximum likelihood estimation produces the same results, and so does a principal

component analysis. MLE estimates reject the null hypothesis that a single factor model is as good as a no factor model. They also reject the null hypothesis that a single factor model is as good as a many factor model. In fact, the chi-square for a 2-factor model (5205) is larger than that for a single-factor model (5023), which suggests that the uniqueness component associated with PS and VA should not be ignored. Further, the difference between these two chi-square values, at 6 degrees of freedom, is itself significant.

Path Analysis:

If the single factor measurement model is correct, the predicted correlation between variables i and j should equal the product of the factor loading (path coefficient or correlation in this case) between variable i and the factor k and between variable j and factor k. We can use a chi-square statistic to test if the observed correlation equals the correlation that would be predicted if the model were true. The results appear in Table 3. There is very little difference between the predicted and actual values. Since

$$\chi^2 = \Sigma \, (\text{Predicted} - \text{Actual})_i^2 \, / \, \text{Predicted}_i, \, i = 1,\ldots,n \text{ pairs},$$

$\chi^2 = .0019$, which is clearly not significant at n-1=14 degrees of freedom. We fail to reject the null hypothesis that the observed correlations differ from those predicted by the measurement model. Overall, this suggests strong evidence in favor of the argument that the WGI variables do not measure distinct concepts. On the other hand, the pairs that contribute most to this small χ^2 are those that involve VA and PS.

Confirmatory Factor Analysis:

Confirmatory Factor Analysis is similar to the informal path analysis model, but it uses MLE procedures to compare observed and predicted values, and explicitly allows for estimates of error. In CFA, each of the 6 WGI variables is regarded as a standardized endogenous variable written as a linear function of a latent exogenous variable and an error term. The error terms are independent of each other, and independent of the latent factor. Because the manifest variables are standardized, the variance of the path coefficient plus the error variance must sum to one. The MLE statistics provide estimates

of the impact of the hypothetical systematic impact of the latent factor on the observed variables, relative to that of the error terms.

Table 4 shows the results. Every path from the one latent variable to each manifest WGI variable is clearly significant substantively and statistically. The path coefficients range from .74 to .96, and no t-statistic is less than 24. This is consistent with the results from the simple path analysis results. However, the CFA procedure explicitly estimates the variance of the error terms. The estimates of the variance of the stochastic terms all have a small standard error. Most importantly, the error variance for PS and VA (.44 and .45, respectively) are nearly as large as the variance of the systematic impact of the latent factor (.56 and .55, respectively).

The Causal Model

It is common to hypothesize that the root causes of corruption are an absence of VA, GE, RQ, RL, and/or PS (Ades and Di Tella, 1999; Bohara, Mitchell and Mittendorf, 2004; Xin and Rudel, 2004; Broadman and Recanatini, 2000; Geddes and Neto, 1999; Johnson, Kaufman, and Zoido-Lobaton, 1998; Kang, 2002; Schleifer and Vishny, 1993). CFA permits a test of this (just identified) hypothesis. In this model, CC is endogenous, and the other WGI variables are exogenous, as is an error term for CC. The error term and the exogenous WGI variables are uncorrelated, but the exogenous WGI variables are correlated with one another. All manifest variables are standardized. The goodness of fit of such a model is quite high; the explained variance is .85. Table 5a reports the estimated OLS coefficients for the causal model. Only two of the coefficients are both significant and have the expected positive sign. According to the results, GE and RL improve CC, as they are expected to do, but RQ reduces it, and, quite unexpected from the strong theoretical expectations from previous research, neither VA nor PS significantly raise CC.

If we take this approach as a serious causal claim, the estimates in Table 5a are unlikely to be unbiased or efficient. First, there is always the possibility of unobserved heterogeneity among nations and years. Adding fixed effects for nations and years reduces (but does not eliminate) this possibility. Further, the data are a panel, and yearly observations within a nation are unlikely to be independent, and the variance of errors

between nations is unlikely to be constant. Thus, estimates from a GLS model, with fixed effects for nation and with each error term corrected for yearly autocorrelation within panels and heteroscedasticity between panels would be more credible than the OLS estimates in Table 5a.

Table 5b shows the FGLS results. Clearly the model fit is very good; yet most of the results in Table 5a do not change, even with the numerous additional control variables. GE and RL remain significant and positive. The puzzling significant negative coefficient for RL disappears, and the indicator becomes insignificant. VA becomes significant and positive in this version of the causal model, but PS remains insignificant.
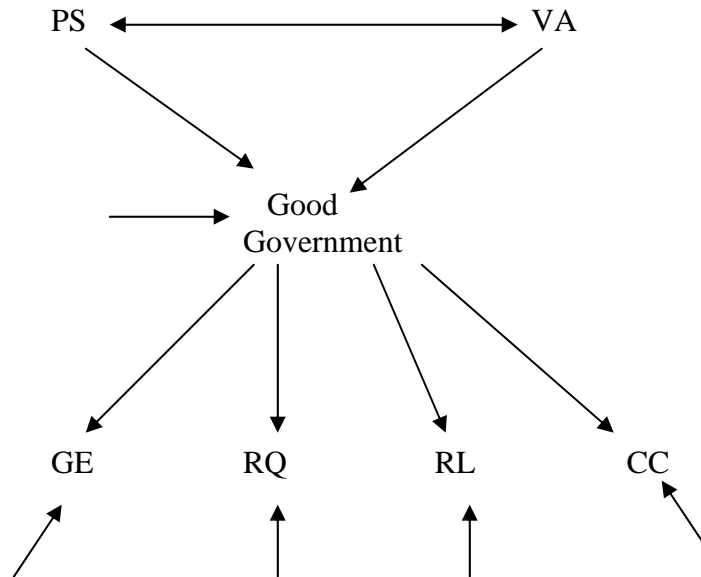
The Mixed Model

Overall, the results so far make all models look quite good; they also suggest that PS (and maybe VA) may not be part of the general concept of good government. Despite the insignificant causal estimates, it is still possible that they do have a direct impact on good government if that concept was properly measured. This suggests a model that includes VA and PS as manifest, exogenous variables that cause good government; good government is measured as a latent variable, represented by GE, RQ, RL, and CC as endogenous variables. Good government is both endogenous to VA and PS, but exogenous to each of its indicators. Each endogenous variable is also affected by an independent error term. The manifest exogenous variables are correlated, but they are all independent of the stochastic terms. Figure 3 shows the mixed model (causal and measurement) and Table 6 shows the results. The results in that Table continue to uphold yet another model. In this model, the evidence suggests that the data conform very well to a model in which PS and VA are exogenous causes of a general concept, which we label "good government," measured by four indicators, GE, RQ, RL, and CC.

CFA goodness of fit indices are often used to test whether one model is better than another, when there are competing theories about how variables are related to one another. In general, these indices are not particularly reliable. Just as with an R-square, good models can have low or high R-square values. However, it is instructive to note that all of the CFA models presented here (a measurement model, a causal model, and a

mixed model) all show goodness-of-fit measures of about .90. This means that all the models are about the same in terms of goodness of fit, and all are about equally good.

Figure 3: Mixed causal and measurement model

Path Analysis of Mixed Model:

Like the CFA measurement model, the mixed model can also be tested using path analysis. Since the model is over-identified, it can be tested. That is, it is possible to compare the correlation between variables that are predicted by the model to have no direct relation (that is, a partial regression parameter of zero) with the observed estimate. In practice, this test is conducted using standardized estimates. Figure 4 replicates Figure 3, adding the standardized estimates of the path coefficient for each direct link. Table 7 reports the path equations that follow if the model is true, and Table 8 compares the predicted and actual correlations. The chi-square test once again summarizes the comparison. With χ^2 (13)= .002, there is a very good correspondence between the data and the mixed model.

Figure 4: Path coefficients: Mixed causal and measurement model
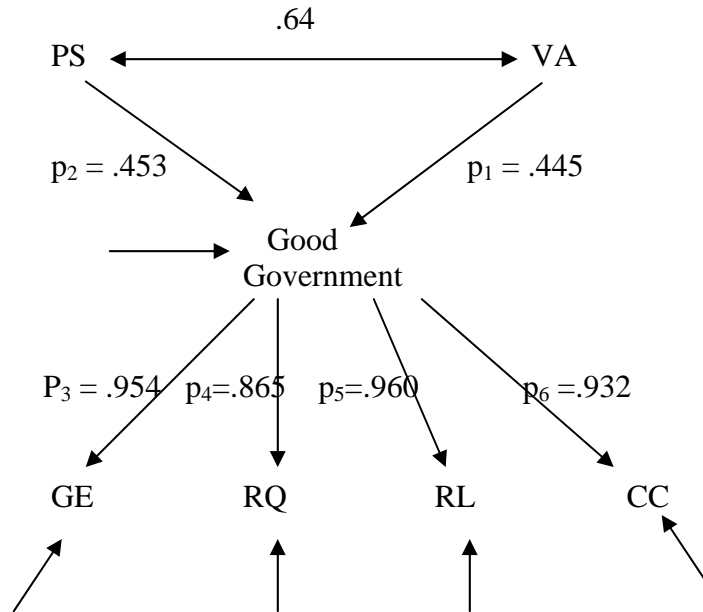
Conclusion

We draw several conclusions from these results. First, the indicators utterly fail to distinguish between the causal, measurement, and mixed models. Variables that uphold all three models equally well probably are themselves not measuring what each claims to measure. It is hard to use these indicators to uphold a causal story when a measurement model, which implies no direct causal links (including two-way links) among the indicators, fits the data just as well. It is difficult to claim that each WGI is distinct. Some clearly represent a more general concept; it is difficult to claim that variables have meaning when they are all consistent with very different theoretical representations. Overall, the results are consistent with the proposition that the separate WGI indicators, because they are, by definition, overlapping, if not equivalent, are tautological.
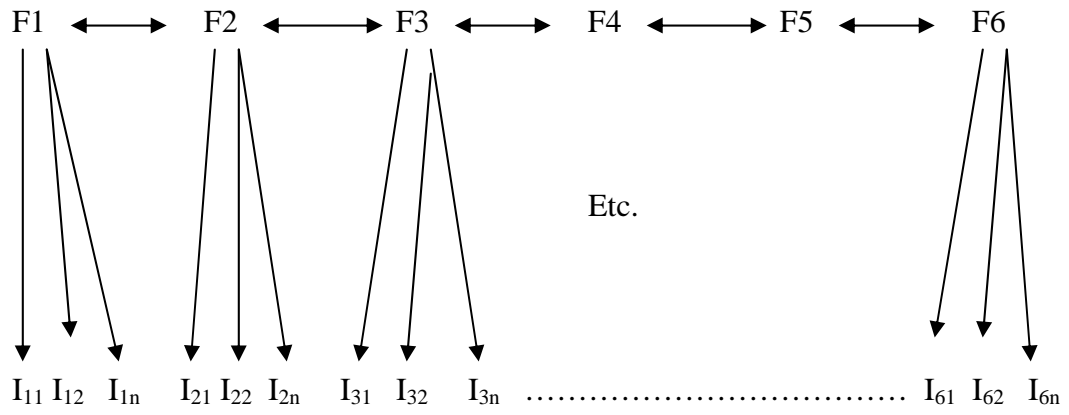
Second, in the data we use, each WGI variable is represented as a "manifest," observed indicator. Yet, in reality, none of the six indicators is a manifest variable. Each WGI indicator is a constructed index comprised of numerous other manifest (i.e.,

measured) variables. The claim is that each WGI measure is conceptually distinct from the other, but this is an untested claim of construct validity. The claim should be tested by subjecting the entire data base of original country-by-year indicators for each of the 6 WGI indexes used in this paper to a CFA in which the null hypothesis is that each measured variable within each of the 6 WGI indexes is a function of only one latent variable, and not of any other variables. Figure 5 represents the model that needs to be tested. Since many of the original indicators are not public information and have not been released, this is not possible to do.

Third, there is no support for some specific "hypotheses" about how indexes are conceptually related. The WGI authors' groupings of VA and PS (the process by which governments are selected, monitored and replaced), RQ and GE (the capacity of the government to effectively, formulate and implement sound policies), and CC and RL (the respect of citizens and the state for the institutions that govern economic and social interactions among them) find no support in the data. Similarly, the MCC's categories of "ruling justly" (including RL, CC, VA, and GE) and "economic freedom" (RQ) find no support in the data. Instead, a single factor explains nearly all of the variation in the six WGI indexes.

Fourth, the results invalidate some practices in the research literature and by aid agency staff. Researchers generally treat the six WGI indexes as measuring distinct concepts, and treat the TI index as a corruption measure, not as a broader indicator of quality-of-governance perceptions. Our results in contrast support the choice of several researchers who have averaged together the six WGI indexes in their analyses into a single broader index, because their high inter-correlations suggested they were not "genuinely measuring different dimensions of governance within each country" (Al-Marhubi, 2004: 396; also see Bjornskov, 2006: 26 and Easterly, 2002). Similarly, in diagnosing governance-related strengths and weaknesses in particular countries, aid agency staff are reading too much into a country's relative scores on CC, RL, VA, RQ, PS and GE.

Figure 5: A complete measurement model

$$F1 \longleftrightarrow F2 \longleftrightarrow F3 \longleftrightarrow F4 \longleftrightarrow F5 \longleftrightarrow F6$$

Etc.

I_{11} I_{12} I_{1n} I_{21} I_{22} I_{2n} I_{31} I_{32} I_{3n} …………………………… I_{61} I_{62} I_{6n}

References

Ades, Alberto and Rafael Di Tella (1999). "Rents, Competition, and Corruption."
American Economic Review 89 (September): 982-993.

Alence, R. (2004). 'Political institutions and developmental governance in sub-Saharan
Africa." *The Journal of Modern African Studies*; June 2004, 42 (2): 163.

Al-Marhubi, F. (2004). "The Determinants of Governance: A Cross-Country Analysis."
Contemporary Economic Policy, Volume 22 (3): 394-406

Bjornksov, Christian (2006). "The Multiple Facets of Social Capital." *European Journal
of Political Economy* 22: 22-40.

Bohara, Alok K., Neil J. Mitchell, and Carl E. Mittendorff (2004). "Compound
Democracy and the Control of Corruption: A Cross-Country Investigation."
Policy Studies Journal 32 (4): 481-498.

Borrmann, A., Busse, M., & Neuhaus, S. (2006). "Institutional Quality and the Gains
from Trade." *KYKLOS*, Vol. 59 (3): 345–368.

Broadman, Harry G. and Francesca Recanatini (2000). "Seeds of Corruption: Do Market
Institutions Matter?" World Bank Policy Research Working Paper 2368 (June).

Brewer, Gene A., Yujin Choi and Richard M. Walker (2007). "Accountability,
Corruption and Government Effectiveness in Asia: An Exploration of World
Bank Governance Indicators." *International Public Management Review* 8 (2).
Available at http://www.ipmr.net.

Damania, Richard; Per G. Fredriksson and Muthukumara Mani (2004). "The persistence
of corruption and regulatory compliance failures: theory and evidence." *Public
Choice* 121: 363-90.

Easterly, William (2002). "Evaluating Aid Performance of Donors." Center for Global
Development (http://www.cgdev.org/doc/CDI/Easterly_Aid_Component.pdf).

Geddes, Barbara and Artur Ribeiro Neto (1999). "Institutional Source of Corruption in
Brazil." In *Corruption and Political Reform in Brazil*, ed. Keith Rosenn and
Downs Rosenn, 21-48. Richard Coral Gables, FL: North-South Center Press.

Islam, Roumeen (2006). "Does more transparency go along with better governance?"
Economics and Politics 18(2): 121-167.

Johnson, Simon, Daniel Kaufmann, and Pablo Zoido-Lobaton (1998). "Regulatory

Discretion and the Unofficial Economy." *American Economic Review* 88(2): 387-392.

Kang, David C. 2002. *Crony Capitalism -Corruption and Development in South Korea and the Philippines*. New York: Cambridge University Press.

Kaufmann, Daniel and Kraay, Aart (2002). "Growth without Governance." *Economica* 3(1): 169–229.

Kaufmann, Daniel; Aart Kraay, and Pablo Zoido-Lobaton (1999). "Aggregating Governance Indicators." World Bank Policy Research Working Paper 2195.

Kaufmann, Daniel, Léautier, Frannie, and Mastruzzi, Massimo (2004). "Governance and the City: An Empirical Exploration into Global Determinants of Urban Performance." Retrieved on October 15, 2007 from http://web.worldbank.org/WBSITE/EXTERNAL/WBI/EXTWBIGOVANTCOR/0,,contentMDK:20725248~menuPK:1976990~pagePK:64168445~piPK:64168309~theSitePK:1740530,00.html

Kurtz, M., & Schrank, A. (2007). "Growth and governance: Models, measures, and mechanisms." *The Journal of Politics*, 69 (2): 538-554.

Lambsdorff, Johann G.(1998). "Transparency International 1998 Corruption Perceptions Index: Framework Document" (http://www.icgg.org/downloads/FD1998.pdf).

Lederman, D., Loayza, N., & Soares, R. (2005). "Accountability and corruption: Political institutions matter." *Economics and Politics* 17 (1): 1-35.

Ritzen, J., Easterly, W., & Woolcock, M. (2000). On 'Good' Politicians and 'Bad' Policies: Social Cohesion, Institutions, and Growth." World Bank Policy Research Working Paper 2248. Washington, D.C.: World Bank.

Sandholtz, Wayne and William Koetzle (2000). "Accounting for Corruption: Economic Structure, Democracy and Trade." *International Studies Quarterly* 44: 31-50.

Shleifer, Andrei and Robert W Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108 (August): 599-617.

Tavits. M. (2007). "Clarity of responsibility and corruption." *American Journal of Political Science* 51(1): 218-229.

Thomas, M.A. (2007). "What do the Worldwide Governance Indicators Measure?" Johns Hopkins University, School of Advanced International Studies, Washington DC.

You, J., & Khagram, S. (2005). "A comparative study of inequality and corruption."
 American Sociological Review 70 (1): 136-157.

Xin, Xiaohui and Thomas K. Rudel (2004). "The Context for Political Corruption: A
 Cross-National Analysis." *Social Science Quarterly* 85 (2), June: 294-309

Table 1: Bivariate Correlations, 6 World Governance Indicators (Country data over 5 years) (N=781)

	VA	PS	GE	RQ	RL	CC
VA	1.00000					
PS	0.6411	1.0000				
GE	0.6920	0.6889	1.0000			
RQ	0.7473	0.6167	0.8564	1.0000		
RL	0.7061	0.7354	0.9099	0.8176	1.0000	
CC	0.6633	0.6861	0.8928	0.7682	0.9079	1.0000

Table 2: Principal component factor analysis of 6 WGI variables (Country data over 5 years) (N=781)

(principal factors; 3 factors retained)

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	4.57990	4.44131	0.9950	0.9950
2	0.13859	0.06191	0.0301	1.0252
3	0.07668	0.11598	0.0167	1.0418
4	-0.03930	0.00249	-0.0085	1.0333
5	-0.04179	0.06958	-0.0091	1.0242
6	-0.11137	.	-0.0242	1.0000

Factor Loadings

Variable	1	2	3	Uniqueness
VA	0.77535	0.23628	0.04939	0.34056
PS	0.75377	0.05442	0.18978	0.39284
GE	0.94535	-0.09004	-0.11225	0.08560
RQ	0.87895	0.15639	-0.15259	0.17971
RL	0.95153	-0.11050	0.04516	0.08034
CC	0.91592	-0.18714	0.01738	0.12577

Table 3: Measurement Model: Predicted and Actual Pairwise Correlations, WGI

variables

Pair	Predicted	Actual
VA, PS	.584	.641
VA, GE	.732	.692
VA, RQ	.681	.747
VA, RL	.738	.706
VA, CC	.710	.663
PS, GE	.712	.689
PS, RQ	.663	.617
PS, RL	.718	.735
PS, CC	.691	.686
GE, RQ	.831	.856
GE, RL	.897	.910
GE, CC	.866	.893
RQ, RL	.837	.818
RQ, CC	.805	.768
RL, CC	.872	.908

Table 4: CFA, One-Latent Factor

Manifest Variable	Path Coefficient	t-statistic	Error variance	t-statistic
VA	.744	24.20	.446	19.02
PS	.745	24.26	.446	19.02
GE	.953	35.88	.091	13.55
RQ	.866	30.37	.250	17.97
RL	.960	36.38	.078	12.42
CC	.932	34.44	.131	15.71

Table 5: Causal Model, The impact of 5 exogenous WGI variables on CC

Panel a: OLS estimates (N=781)

Independent Variable	Path Coefficient	t-statistic
GE	.432	11.45
RQ	-.082	-2.78
RL	.55	15.18
VA	.025	1.14
PS	.018	0.83
R^2	.852	

Panel b: FLGS estimates (Fixed effects for nation and state not shown; corrected for within panel ar1 and between panel heteroscedasticity; N = 704)

Independent Variable	Coefficient	z-statistic
GE	.127	6.06
RQ	.005	0.34
RL	.427	14.96
VA	.136	7.11
PS	.021	1.83
Wald Chi2 (134)	96159	

Table 6: Mixed model CFA estimates: Good Government as latent independent and dependent variable.

Manifest variable equation estimates

	Coefficient for Good Government	t-statistic	R-Sq.
Dependent variable			
GE	.98	90.8	.91
RQ	.89	54.9	.75
RL	.99	94.8	.92
CC	.96	78.5	.87
Latent variable equation estimates			.67
Independent Variables			
VA	.43	16.1	
PS	.43	16.3	

Table 7: Path Equations, Mixed causal and measurement model (stochastic terms
deleted for convenience)

$F1 = p_1\ VA + p_2\ PS$

$GE = p_3\ F1$

$RQ = p_4\ F1$

$RL =\ p_5\ F1$

$CC = p_6\ F1$

$r_{VA,GE}\quad = p_3\ p_1 + p_3\ p_2\quad r_{VA,PS}$

$r_{PS,GE}\quad = p_3\ p_1\ r_{VA,PS}\ +\quad p_3\ p_2$

$r_{VA,RQ}\quad = p_4\ p_1 + p_4\ p_2\quad r_{VA,PS}$

$r_{PS,RQ}\quad = p_4\ p_1\ r_{VA,PS}\ +\quad p_4\ p_2$

$r_{VA,RL}\quad = p_5\ p_1 + p_5\ p_2\quad r_{VA,PS}$

$r_{PS,RL}\quad = p_5\ p_1\ r_{VA,PS}\ +\quad p_5\ p_2$

$r_{VA,CC}\quad = p_6\ p_1 + p_6\ p_2\quad r_{VA,PS}$

$r_{PS,CC}\quad = p_6\ p_1\ r_{VA,PS}\ +\quad p_6\ p_2$

$r_{GE,RQ}\quad = p_3\ p_4$

$r_{GE,RL}\quad = p_3\ p_5$

$r_{GE,CC}\quad = p_3\ p_6$

$r_{RQ,RL}\quad = p_4\ p_5$

$r_{RQ,CC}\quad = p_4\ p_6$

$r_{RL,CC}\quad = p_5\ p_6$

Table 8: Mixed Causal and Measurement Model: Predicted and Actual Pairwise

Correlations, WGI variables

Pair	Predicted	Actual
VA, GE	.702	.685
PS, GE	.704	.682
VA, RQ	.636	.747
PS, RQ	.638	.617
VA, RL	.705	.708
PS, RL	.709	.737
VA, CC	.685	.663
PS, CC	.687	.686
GE, RQ	.825	.856
GE, RL	.916	.905
GE, CC	.889	.893
RQ, RL	.830	.818
RQ, CC	.806	.768
RL, CC	.895	.908