Economics Bulletin

Volume 39, Issue 1

An interval variables approach to address measurement uncertainty in governance indicators

Carlo Drago
University of Rome "Niccolò Cusano" and NCI University in London

Roberto Ricciuti
University of Verona and CESifo

Abstract

Many variables in governance are measured with uncertainty. This paper addresses this problem, showing that interval variables are a suitable way to handle it, providing an application in corporate governance. We build two constructs, one for Investor protection and the other for Constraints on shareholders based on the original dataset by La Porta et al. (1998) and we find that for very low levels of investor protection, constraints are a suitable way to provide some form of safeguard. We also provide evidence for the theoretical claim that investor protection and constraints on shareholders work as substitutes under specific circumstances.

We thank two reviewers for useful comments. Any remaining errors are ours.
Citation: Carlo Drago and Roberto Ricciuti, (2019) "An interval variables approach to address measurement uncertainty in governance indicators", *Economics Bulletin*, Volume 39, Issue 1, pages 626-635
Contact: Carlo Drago - md3687@mclink.it, Roberto Ricciuti - roberto.ricciuti@fulbrightmail.org.
Submitted: September 05, 2018. **Published:** March 16, 2019.

1. Introduction

Measuring governance has been a major endeavor in the last twenty years both at the macro (evaluating institutions and policies) and micro level (companies). This effort has been both methodological and empirical, trying to understand what can be measured and its nature (*de facto* or *de jure*), but also suffered from a number of issues that have been highlighted in the literature. Landmann and Hauserman (2003) list these issues as: 1) Validity (the indicator measures the object in question); 2) Reliability (ability to reproduce the indicator using the same coding rules and the same hardware support); 3) Measurement bias (the question of whether there is a systematic measurement error; 4) Lack of transparency in the production of the indicator; 5) Representativeness (sample type for survey data); 6) Variance truncation; 7) Information bias (information sources used); 8) Aggregation problem (how complicated or inconsistent scores aggregation rules).[1]

Most of the debate has been centered on the lack of a theoretical model on which these indices are grounded, and on their subjective/objective nature (Kerim et al., 2015). Subjective data are deemed to be imprecise, biased (possibly by the political views or by the role that people surveyed have), and driven by factors other than governance itself, such as the level of development of a country. It is fair to say that the distinction between "subjective" and "objective" data is often less straightforward than it might seem. "Objective" measures of governance are usually based on the coding of formal laws and regulations, they require judgment as the net effect of potentially conflicting rules.

All these issues point to the fact that these indicators are constructed with uncertainty (Kaufmann et al., 2007; OECD, 2008). World Governance Indicators, for example, are provided giving a point estimate and an interval to measure uncertainty.[2] The aim of this paper is to introduce some tools from statistics, namely interval variables and interval regression, that address the problem of uncertainty in governance indicators, and to provide a simple application. The intuition behind it is as follows: suppose we want to measure a governance feature, and we already have some variables in the literature that partially capture that issue; we propose to take all possible combinations of these variables to provide a construct[3] characterizing the feature we are interested in. From each combination we extract a statistical value that captures the commonality among the variables, and then from the different values obtained we build an interval of values between the minimum and the maximum. The uncertainty in measurement may be high or low, depending on the size of this interval.

Interval variables belong to the Symbolic Data Analysis paradigm (Billard and Diday, 2006). This approach represents a transition from "individual observations" described by standard variables of numerical and categorical values to "higher level observations" described by variables of symbolic values taking care of their internal variation (intervals, probability distributions, sets of categories or numbers, random variables, etc.) which cannot be treated as numbers. While individual observations are a player, a stock, etc., examples of higher level observations are classes (a player subset, a subset of funds, …), categories (American funds, European funds, …), concepts: volatile American funds. Governance variables belong to the latter type of observations, therefore interval variables may be a better option compared with a

[1] Voigt (2013) provides a throughout discussion on what and how to measure institutions.

[2] Kaufmann et al. (2011) argue that the presence of margins of error is not a consequence of the use of subjective data to measure governance but is due to the fact that available data are imperfect proxies for the concepts they are trying to measure. However, we need to point out that the way in which they are calculated is pretty different from ours, and we also aim to use these intervals to run regressions between these constructs in order to take into account this uncertainty when establishing empirical regularities.

[3] A construct is the abstract idea that one wishes to measure using survey questions. Complex constructs contain multiple dimensions or facets that are bound together by some commonality that, as a whole, compose the construct.

classical way to measure concepts in governance that use single observations. In this way data are not forced into single values and therefore can provide more information, resulting in more reliable estimates. We deem this a very important feature of the method, since we are trying to capture latent variables which are not directly observed but are rather inferred from other variables that are observed and directly measured, with the inherent degree of uncertainty that we have highlighted.

The paper is organized as follows: after introducing interval variables (section 2), we apply this method in section 3 to the well-established dataset on law and finance provided by La Porta et al. (1998). We provide two constructs, one for investor protection and another for constraints on minority shareholders, and we show that for the subsample of low investor protection countries, there is a negative relationship between them, which we interpret as constraints substituting for investor protection. Section 4 concludes.

2. Interval variables

Since using the mean, the median or a function to aggregate these data clearly leads to an information loss, we consider an approach that explicitly takes into account the variation in data, such as interval data (Billard and Diday 2003).

Starting with c composite indicators with $c = 1, \dots C$ and k constructs with $k = 1, \dots . K$, we build the interval variable by considering different representations of the single composite indicator:

$$I[Y]^c = [\underline{Y_k^c}, \overline{Y_k^c}] \qquad (1)$$

where $\underline{Y_k^c}$ and $\overline{Y_k^c}$ are the lower and the upper bounds of the interval.

There are various statistical approaches to build composite indicators, we consider the outcomes from the Principal Component Analysis (PCA thereafter). The advantage of PCA is that we do not need to consider a specific weighting of the various constructs that can lead to very different results. We start from the data matrix and we perform each PCA, then the results from the PCA analysis are considered.

Before building a unique construct, we need to test its internal consistency whose measure is based on the correlations of the different variables adopted. In other words, we need to detect whether different variables can lead to the construction of a single unidimensional construct. To do so, for each group of variables we calculate the Cronbach alpha (Cronbach, 1951):

$$\alpha_c = \left(\frac{Q}{Q-1}\right)\frac{\sum_{i \neq j} cov(x_i, x_j)}{var(x_0)} \quad c = 1 \dots C \; ; \; i, j = 1 \dots Q \qquad (2)$$

Where c are the constructs, Q is the number of the different indicators considered at the start of the measurement process, x_i and x_j are two indicators considered in the construction of the latent variable, and x_0 is the sum of all the indicators considered in building the construct. A group of variables is considered if the Cronbach alpha is around 0.80. This number is established in literature as the relevant one in order to obtain a single principal component in the data matrix (Tavakol and Dennick, 2011). The groups of variables used in the measurement of the construct that show a low Cronbach alpha are discarded. The constructs obtained in this way are considered for the interval of values.

From each PCA we extract the first component, since this is the most informative to measure the latent construct. The other components (eigenvalue criterion smaller than one) are

less informative than the original variables (Gherghi and Lauro, 2004). Therefore, for each composite indicator *c* we will consider all the possible specifications *k* and so we obtain a different outcome for the composite indicator Y_k^c .

2.1 The interval variables characteristics of the constructs

The different composite indicators represent the uncertainty in measuring the complex concept. We can usefully represent this complexity as an interval variable, which can be characterized by its lower and upper bounds, while other characteristics of the intervals of the constructs are the centers and the radii. Being *c* the constructs considered and *k* the interval variables obtained, we can have:

$$Y_{center,k}^{c} = \frac{1}{2}(\underline{Y_k^c} + \overline{Y_k^c}) \qquad (3)$$

Where $\underline{Y_k^c}$ and $\overline{Y_k^c}$ are the lower and the upper bound of the constructs, respectively. The center can be defined as the central measure of the interval considered, the measure to compare different interval variables *k*. Moreover, it is also possible to take into account a measure of the variation between the two bounds, the radius of the interval and the upper and the lower bounds:

$$\mathrm{Y}_{\mathrm{radius,k}}^{\mathrm{c}} = \frac{1}{2}(\underline{\mathrm{Y}_{\mathrm{k}}^{\mathrm{c}}} - \overline{\mathrm{Y}_{\mathrm{k}}^{\mathrm{c}}}) \qquad (4)$$

2.2 Interval regression

We can now consider some basic statistical methods to analyze the interval indicators we have previously constructed. The mean of the interval variable is:

$$M = \left[\frac{1}{n}\sum_{i=1}^{n}\underline{Y}\,;\frac{1}{n}\sum_{i=1}^{n}\overline{Y}\right] = [\underline{M};\overline{M}] \qquad (5)$$

where $\underline{Y}$ is the lower bound of the interval variable measured for the different observations and $\overline{Y}$ is the upper bound. The interval variables are obtained for each observation (for example different measurements of the latent variable for each country). Here the mean *M* is the average of the intervals, with $\underline{M}$ the lower bound and $\overline{M}$ is the upper bound.

The linear regression model is the most frequently used form in interval regression analysis for expressing the relationship between one or more explanatory variables and response. For the sake of simplicity, the case of simple linear regression model involving a single independent variable is considered, which can be easily generalized to the case of multiple inputs, although computationally more complex.

The objective of the interval regression is to determine a functional linear relationship:

$$Y = A_0 \oplus A_1 \otimes X \qquad (6)$$

where *Y* is the interval model output, *X* the interval model input and A_0 and A_1 the interval parameters, i.e. unknowns to be estimated from interval data.

The coefficients of the model can be estimated by applying the classical model to the mid-point of the intervals (Billard 2008, Billard and Diday 2000). The estimation of the parameters may be obtained by an adaptation of the solution obtained by the Least Square

estimation method for the classical linear model, where relevant definitions of variance and covariance are used.

3. An application to the law and finance literature

La Porta et al. (1998) examine legal rules covering protection of corporate shareholders and creditors, the origin of these rules, and the quality of their enforcement in 49 countries. They show that common law countries generally have the strongest legal protections of investors, with German and Scandinavian civil law countries located in the middle, and French civil law countries at the bottom.

Do these countries have other, *substitute* mechanisms of corporate governance? The literature has highlighted a number of potential candidates: strong enforcement of laws, mandatory standards of retention and distribution of capital to investors, and high ownership concentration. La Porta et al. (1998) find a negative correlation between concentration of ownership and the quality of legal protection of investors. We choose this dataset for his prominence in the law and economics literature but, as we argue in the Conclusions, this approach can be applied to other datasets.

The aim of the application is to analyze the relationship between investor protection and constraints on shareholders. Both are measured with uncertainty; therefore, we use intervals data in order to capture this aspect. The first step in the analysis is considering the dataset of La Porta et al. (1998) to measure investor protection and constraints on shareholders. We consider two sets of variables and apply to the entire set of variables the Cronbach alpha, obtaining both constructs from the initial indicators. Construct 1 is built from rule of law, efficiency of the judicial system, repudiation of contracts by government, expropriation risk, accounting standards. Construct 2 comes from extraordinary shareholders' meeting votes, mandatory dividend, ownership of 10 largest private firms and legal reserves. The Appendix reports the definitions and the sources of the original variables. Then we perform a principal component analysis on the set of the variables considered by applying the criterion of the eigenvalue value for each group of variables. Tables 1 and 2 show the intervals, the centers and the radii of each construct.

TABLE 1 - Construct 1: Investor protection

Country	Lower Bound	Upper Bound	Center	Radius
Argentina	-2.16	-1.60	-1.88	0.28
Australia	1.45	1.86	1.66	0.20
Austria	1.79	2.16	1.97	0.18
Belgium	1.73	2.10	1.91	0.19
Brazil	-1.09	-0.70	-0.89	0.20
Canada	1.50	1.91	1.71	0.20
Chile	-0.59	-0.28	-0.43	0.15
Colombia	-1.63	-0.72	-1.18	0.46
Denmark	1.81	2.17	1.99	0.18
Ecuador	-1.75	-0.98	-1.36	0.39
Egypt	-1.82	-1.44	-1.63	0.19
Finland	1.76	2.12	1.94	0.18
France	1.13	1.59	1.36	0.23
Germany	1.61	2.02	1.81	0.21
Greece	-0.87	-0.65	-0.76	0.11
Hong Kong	0.79	1.35	1.07	0.28
India	-1.18	-0.53	-0.86	0.32
Indonesia	-2.52	-1.44	-1.98	0.54
Ireland	0.97	1.37	1.17	0.20

Israel	-0.38	0.66	0.14	0.52
Italy	0.60	1.33	0.97	0.36
Japan	1.72	2.08	1.90	0.18
Jordan	-2.17	-1.08	-1.62	0.55
Kenya	-1.95	-1.48	-1.72	0.23
Malaysia	-0.10	0.29	0.10	0.20
Mexico	-1.21	-0.95	-1.08	0.13
Netherlands	1.94	2.29	2.11	0.18
New Zealand	1.81	2.17	1.99	0.18
Nigeria	-2.97	-2.10	-2.53	0.43
Norway	2.03	2.36	2.19	0.17
Pakistan	-2.93	-2.49	-2.71	0.22
Peru	-2.83	-2.19	-2.51	0.32
Philippines	-3.20	-2.65	-2.93	0.27
Portugal	0.11	1.04	0.58	0.46
Singapore	1.26	1.63	1.45	0.19
South Africa	-1.44	-0.98	-1.21	0.23
South Korea	-0.68	0.10	-0.29	0.39
Spain	0.13	1.02	0.58	0.45
Sri Lanka	-2.59	-1.74	-2.16	0.42
Sweden	1.80	2.16	1.98	0.18
Switzerland	2.07	2.47	2.27	0.20
Taiwan	0.56	1.28	0.92	0.36
Thailand	-1.53	-0.37	-0.95	0.58
Turkey	-1.94	-1.29	-1.61	0.32
UK	1.64	2.00	1.82	0.18
US	1.82	2.18	2.00	0.18
Uruguay	-1.29	-0.83	-1.06	0.23
Venezuela	-1.18	-0.85	-1.02	0.17
Zimbabwe	-2.43	-1.63	-2.03	0.40

TABLE 2 - Construct 2: Constraints

Country	Lower Bound	Upper Bound	Center	Radius
Argentina	0.12	0.44	0.28	0.16
Australia	-1.67	-1.47	-1.57	0.10
Austria	-0.18	0.12	-0.03	0.15
Belgium	0.12	0.55	0.33	0.21
Brazil	1.87	2.20	2.04	0.17
Canada	-1.67	-1.47	-1.57	0.10
Chile	1.41	1.50	1.45	0.04
Colombia	3.17	3.72	3.44	0.28
Denmark	0.50	0.60	0.55	0.05
Ecuador	3.17	3.72	3.44	0.28
Egypt	1.24	1.40	1.32	0.08
Finland	-0.23	-0.20	-0.22	0.01
France	0.06	0.12	0.09	0.03
Germany	-0.18	0.12	-0.03	0.15
Greece	1.73	2.09	1.91	0.18
Hong Kong	-1.47	-1.43	-1.45	0.02
India	-1.47	-1.43	-1.45	0.02
Indonesia	-0.23	-0.20	-0.22	0.01
Ireland	-1.47	-1.43	-1.45	0.02
Israel	-1.47	-1.43	-1.45	0.02

Italy	0.44	0.84	0.64	0.20
Japan	0.17	0.60	0.38	0.22
Jordan	0.60	1.23	0.91	0.31
Kenya	-1.47	-1.43	-1.45	0.02
Malaysia	-1.47	-1.43	-1.45	0.02
Mexico	0.44	1.47	0.95	0.51
Netherlands	-0.23	-0.20	-0.22	0.01
New Zealand	-1.67	-1.47	-1.57	0.10
Nigeria	-1.47	-1.43	-1.45	0.02
Norway	0.36	0.44	0.40	0.04
Pakistan	-1.47	-1.43	-1.45	0.02
Peru	0.44	0.84	0.64	0.20
Philippines	-0.20	-0.18	-0.19	0.01
Portugal	0.12	0.44	0.28	0.16
Singapore	-1.47	-1.43	-1.45	0.02
South Africa	-1.67	-1.47	-1.57	0.10
South Korea	1.00	1.40	1.20	0.20
Spain	0.12	0.44	0.28	0.16
Sri Lanka	-1.47	-1.43	-1.45	0.02
Sweden	0.36	0.44	0.40	0.04
Switzerland	1.24	1.40	1.32	0.08
Taiwan	2.37	3.01	2.69	0.32
Thailand	-1.14	-0.66	-0.90	0.24
Turkey	0.36	0.44	0.40	0.04
UK	-1.47	-1.43	-1.45	0.02
US	-1.47	-1.43	-1.45	0.02
Uruguay	1.15	1.54	1.34	0.20
Venezuela	0.12	0.55	0.33	0.21
Zimbabwe	-1.67	-1.47	-1.57	0.10

The coefficients of the regression line are estimated through the mid-point approach outlined above. More specifically, we regress the interval variables for the investor protection and the constraints on shareholders considering two groups of intervals: higher and lower investor protection (calculated with respect to the mean of the investor protection).[4] We find that for high levels of investor protection the estimated coefficient is not statistically significant, whereas there is a positive and statistically significant relationship for countries with low values of investor protection making constraints work as a substitute for explicit regulations for investor protection.

TABLE 3 - Estimation

	Model 1: Low Protection	Model 2: High Protection
Intercept	-1.668**	1.508***
	(0.137)	(0.137)
Constraints	0.216*	0.095
	(0.087)	(0.122)
Adjusted R^2	0.197	-0.017
F-statistic	6.048**	0.596

[4] Computations are performed in R language with the RSDA package (Rodriguez et al., 2014).

Graphically, each observation is represented by a rectangle, which embodies the uncertainty in measuring each variable, since the value lies between a high and a low value. Figures 1 and 2 report the regression line previously estimated and the observations.

FIGURE 1 – High protection

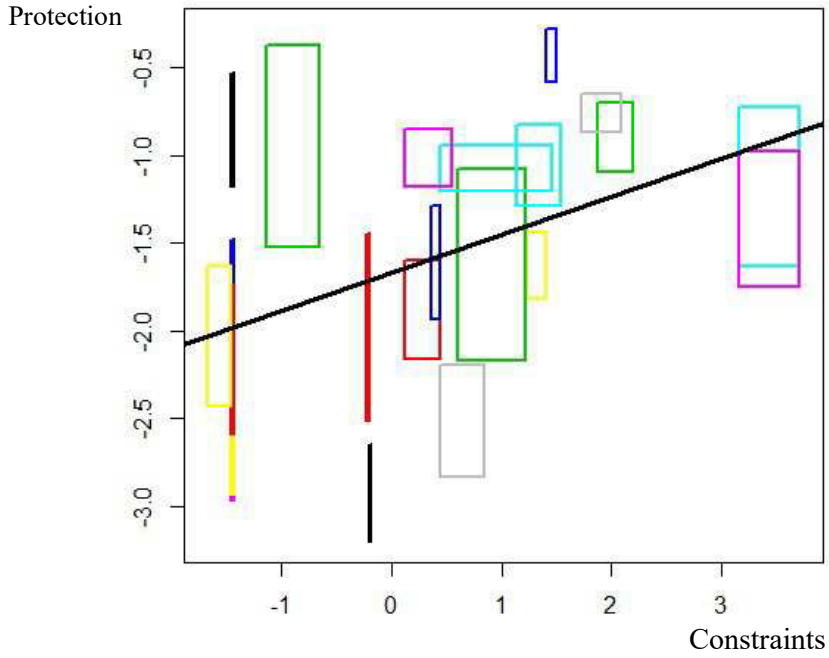
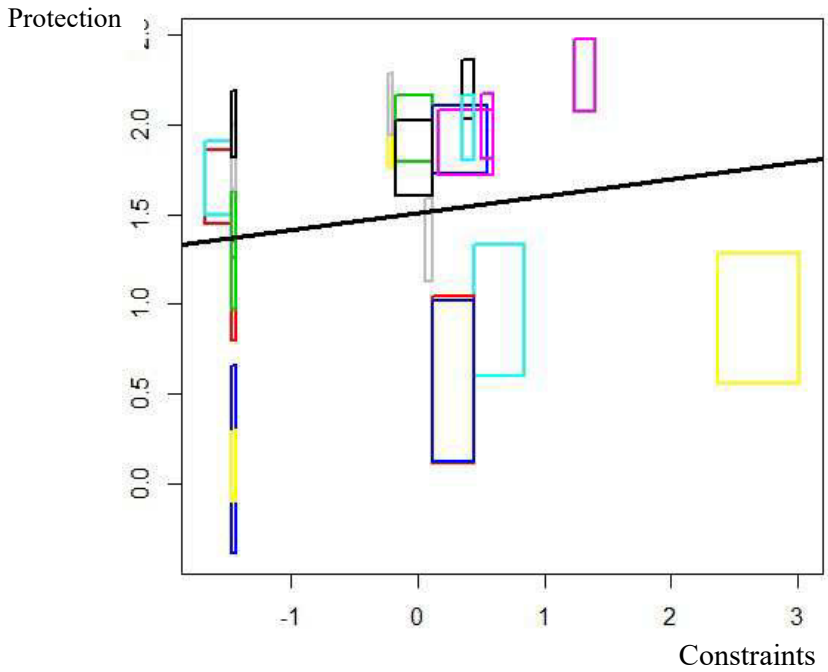


FIGURE 2 – Low protection



4. Conclusions

This paper has suggested the use of interval variables as an effective way to deal with uncertainty in measuring corporate governance variables. This uncertainty may come from a number of sources: lack of a theoretical foundation on how to measure a certain variable, measurement errors in both left- and right-hand side variables, etc. Interval variables depart from the measurement of a single variable and allow building constructs obtained from different variables measuring a similar phenomenon. Interval variables explicitly take uncertainty into account. The center of the interval can be considered as the value that represents more likely the "real" value, and the size of the interval is a measure of the uncertainty.

We apply this methodology the well-known La Porta et al. (1998) dataset on law and finance. We build two constructs (one for investor protection and the other for constraints on shareholders) and we find that in a bivariate interval regression constraint work as a substitute for investor protection in countries in which the latter is small. Clearly, these results are very simple and do not come from a full-size econometric model in which several covariates are included.

We have provided a primer on the use of a statistical tool that may be applied to other datasets (such as World Governance Indicators, the Database of Political Institutions, G-score (Gompers et al., 2003), etc.) and address some important issues in governance measurement and the relation of governance variables with economic, institutional, and social variables.

References

Billard, L. (2008) "Some analyses of interval data" CIT. Journal of Computing and Information Technology, 16(4), 225-233.

Billard, L., and Diday, E. (2006) *Symbolic Data Analysis: conceptual statistics and data mining*. Wiley.

Billard, L., and Diday, E. (2003) "From the statistics of data to the statistics of knowledge: symbolic data analysis" Journal of the American Statistical Association, 98(462), 470-487.

Billard, L., and Diday, E. (2000) "Regression analysis for interval-valued data" In: Kiers, H.A.L, Rasson, J.P., Groenen, P.J.F. and Schader, M. (eds.), *Data Analysis, Classification and Related Methods*, Heidelberg, Springer.

Cronbach, L.J. (1951) "Coefficient alpha and the internal structure of tests" Psychometrika, 16(3), 297-334.

Gherghi, M. and Lauro, C. (2004) *Appunti di analisi dei dati multidimensionali: metodologia ed esempi*. RCE edizioni.

Gompers, P., Ishii, J. and Metrick, A. (2003) "Corporate governance and equity prices" Quarterly Journal of Economics, 118, 107–155.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2011) "The Worldwide Governance Indicators: Methodology and analytical issues" Hague Journal of the Rule of Law, 3(2), 220-246.

Kaufmann, D., Kraay, A., and Mastruzzi, M. (2007) "The Worldwide Governance Indicators project: answering the critics" Research working paper no. WPS 4149. Washington, DC: World Bank.

Kerim, S.A., Rey-Valette, H., Seyte, F., Boccanfuso, D. (2015) "A subjective view of governance indicators" LAMETA DR n°2015-10.

Landmann, T., Hauserman, J. (2003). “Map-making and analysis of the main international initiatives on developing indicators on democracy and good governance”, University of Essex and Eurostat.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. W. (1998) “Law and finance” Journal of Political Economy, 106(6), 1113-1155.

OECD (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide*, Paris.

Rodriguez, R.O. with contributions from Calderon O. and Zuniga R. (2015) RSDA - R to Symbolic Data Analysis. R package version 1.3. http://CRAN.R-project.org/package=RSDA.

Tavakol, M., and Dennick, R. (2011) “Making sense of Cronbach's Alpha” International Journal of Medical Education, 2, 53.

Voigt, S. (2013) “How (not) to measure institutions” Journal of Institutional Economics, 9, 1-26.

Appendix

Components of Construct 1	
Rule of law	Assessment of the law and order tradition in the country produced by the country-risk rating agency *International Country Risk* (ICR). Average of the months of April and October of the monthly index between 1982 and 1995. Scale from 0 to 6, with lower scores for less tradition for law and order.
Efficiency of the judicial system	Assessment of the "efficiency and integrity of the legal environment as it affects business, particularly foreign firms" produced by the country-risk rating agency *Business International Corporation.* It "may be taken to represent investors' assessments of conditions in the country in question". Average between 1980-1983. Scale from 0 to 10, with lower scores lower efficiency levels.
Repudiation of contracts by governments	ICR's assessment of the "risk of a modification in a contract taking the form of a repudiation, postponement, or scaling down" due to "budget cutbacks, indigenization pressure, a change in government, or a change in government economic and social priorities." Average of the months of April and October of the monthly index between 1982 and 1995. Scale from 0 to 10, with lower scores for higher risks.
Risk of expropriation	ICR's assessment of the risk of "outright confiscation" or "forced nationalization". Average of the months of April and October of the monthly index between 1982 and 1995. Scale from 0 to 10, with lower scores for higher risks.
Accounting standards	Index created by examining and rating companies' 1990 annual reports on their inclusion or omission of 90 items. These items fall into 7 categories (general information, income statements, balance sheets, funds flow statement, accounting standards, stock data and special items). From International Accounting and Auditing Trends, Center for International Financial Analysis & Research, Inc.
Components of Construct 2	
Extraordinary Shareholders' Meeting Votes	Percentage of share capital to call an extraordinary shareholders' meeting It is the minimum percentage of ownership of share capital that entitles a shareholder to call for an Extraordinary Shareholders' Meeting.
Mandatory dividend	Equals the percentage of net income that the Company Law or Commercial Code requires firms to distribute as dividends among ordinary stockholders. It takes a value of zero for countries without such restriction.
Ownership of 10 largest private firms	The average percentage of common shares owned by the three largest shareholders in the ten largest non-financial, privately-owned domestic firms in a given country. A firm is considered privately owned if the State is not a known shareholder in it.
Legal reserves	It is the minimum percentage of total share capital mandated by Corporate Law to avoid the dissolution of an existing firm. It takes a value of zero for countries without such restriction.