The Machine Learning Political Indicators Dataset

James T. Bang
St. Ambrose University
518 W. Locust Street
Davenport, IA 52803
Phone: 563-333-6102
Cell: 540-460-5948
Fax: 563-333-6268
Email: BangJamesT@sau.edu

Atin Basu
Virginia Military Institute

Aniruddha Mitra
Bard College

May 21, 2015

Abstract

We introduce a database of institutional variables using machine learning techniques. We perform exploratory factor analysis (EFA) on 32 commonly-cited measures of institutional quality and stability. Our results reduce these 32 variables to 9 theoretically relevant factors: Democracy (based on measures of electoral competition, political dispersion, and the polity 2 index); protest (based on numbers of strikes, riots, and demonstrations); within-regime instability (based on measures of division within the governing coalition); credibility (based on the prevalence of political and economic crises); adverse regime change (based on coups, purges, other sudden changes in the executive branch); flexibility (based on executive and veto player turnover and democracy); regime instability (based on nonviolent changes in the executive and legislature); violence (based on the presence of political violence); and transparency (based on measures of regime durability and the absence of fraud). Relative to current measures, our database has the advantage that the factors we derive are easily reproducible, relatively uncorrelated, robust to missing values, and adaptive to (i.e. "learn" from) the incorporation of new data.
Keywords: Institutions, Political Instability, Factor Analysis, Machine Learning
JEL Codes: C82, D74, O17, P48,

1. Introduction

The importance of institutions has been a centerpiece in explaining empirical differences in economic performance across countries for about the last quarter century, with important early contributions from both economics (North, 1989; Knack and Keefer, 1995) and political science (North and Weingast, 1989; Przeworski and Limongi, 1993; Weingast, 1993; Weingast, 1995). Since these earlier contributions, studies have investigated and found an important role for institutions in explaining a wide variety of phenomena related to performance, including: Institutional volatility and growth (Aisen and Veiga, 2013; Fatás and Mihov, 2013); infrastructure development (Esfahani and Ramirez, 2003); foreign direct investment (Li and Resnick, 2003; Asiedu, 2006; Busse and Hefeker, 2007); trade and trade policy (Morrow and Siverson, 1998; Adsera and Boix, 2002; Milner and Kubota, 2005); migration and the brain drain (Bang and Mitra, 2011; Bang and Mitra, 2013; Mitra, et al. 2014); internal conflict (Easterly, 2001; Bang and Mitra, 2015; Basu et al., 2015) and even the role of institutions in determining the impacts of natural disasters (Kahn, 2005; Cavallo et al., 2013).

Despite the general agreement over the salience of institutions for performance, studies differ greatly over their preferred measure of institutions (Knack and Keefer, 1995; Jong-A-Pin, 2009; Bang and Mitra 2011). While some studies emphasize democracy, others emphasize property rights, transparency, or political stability. Thus, and as noted by Langbein and Knack (2010), there may not be a single underlying concept of "institutions." Moreover, within each broad concept of institutional quality, there may exist several different *structural* institutional characteristics that measure that underlying concept.

This paper aims to clear the water with respect to measuring institutional quality and stability. By considering thirty different measures in an exploratory factor analysis, we are able to construct nine underlying concepts that capture the various dimensions of institutional quality and stability: Democracy, transparency, credibility, civil protest, violence, instability within the regime, nonviolent regime change, adverse regime change, and institutional flexibility. We hypothesize that these dimensions matter differently for economic performance.

To appreciate the possibility of such bifurcations in the effects of institutions, consider the relationship between instability and growth. Certain types of instability (violence and adverse regime changes) may be detrimental to growth because of the uncertainty they inherently introduce risk to economic interactions. Other types of instability (such as institutional flexibility and civil protest) may be benign or even beneficial because they demonstrate the ability of the regime to accept feedback from various constituencies and incorporate a plurality of interests into policy decisions.

The remainder of the paper proceeds with a review of some key contributions to the literature on institutions and different measures of institutional quality in section two; a description of our data sources and methodology in section three; a discussion of the new institutional factors in section four; and concluding remarks and policy implications in section five.

2. Related Literature

A vast and distinct literature documents the importance of institutions for economic development. Although this literature is not the direct scope for the present research, many of the contributions to this literature have been formative in beginning a discourse in the profession about the importance of institutions' for performance, and therefore the importance of measuring institutional quality. Early contributions to this literature focus on the role of institutions in defining property rights (Coase, 1960; Demsetz, 1967), overcoming collective action problems (Olson, 1965; Ostrom, 1990), and economizing on transactions costs (Williamson, 1979; North, 1981). For a comprehensive review of these and many other key contributions, we refer the reader to Lin and Nugent (1995).

Since these foundational contributions, subsequent research has focused on the consequences of poor institutional quality in causing the growth tragedy of Africa (Easterly and Levine, 1997) and the growth collapses in South America and the Middle East (Rodrik, 1999). In these cases, poor performance has been traced to social conflict, which in turn came as the result of weak or unstable institutions. A broader literature has also emerged investigating the effects of inequality and violence on economic performance. In one branch of this literature, Alesina and Perotti (1996), Perotti (1996) and Sala-i-Martin (1997) develop measures of instability, whereas in another, Knack and Keefer (2002) focus on the security of property rights. An important finding of our research is that isolating these dimensions of institutional quality may not be entirely straightforward, and in fact each dimension may depend on multiple structural characteristics of a country's institutional structure.

One reason that measuring institutions may be difficult is that many of the available measures used in most studies of institutional quality are correlated. Most existing studies account for this correlation by either including individual measures separately in analyses of the impacts of institutions (Easterly and Levine, 1997; Grogan and Moers, 2001), or alternatively they construct scalar indices of institutional structure using principle components or logit analysis (Alesina and Perotti, 1996; Perotti, 1996). Note that if multiple dimensions of institutional quality matter, the first method may fail due to omitted variable bias. Yet even if this is not the case, the fact that these measures often in fact capture elements of multiple concepts of institutions exposes these measures to considerable measurement error. However, the second approach entirely ignores the issue of multidimensionality, and therefore fails to recognize which specific attributes of institutional quality matter *most.*

Highlighting the need to account for the multidimensional nature of institutions, Langbein and Knack (2010) perform a confirmatory factor analysis of the World Governance Indicators published by the World Bank. They fail to confirm that these six measures of good governance causally relate to a single latent variable. By contrast, Ghate et al. (2003), Jong-A-Pin (2009) and Bang and Mitra (2011 and 2015) account for multidimensional institutional quality and stability, and find these dimensions to differ in their impacts on growth.

The present approach combines elements of Jong-A-Pin (2009) and Bang and Mitra (2011) in that we combine a long list of measures of institutional quality and stability into an exploratory factor analysis (EFA) to develop a general view of institutions. Not surprisingly, our results mirror these two studies closely, and we find evidence for eight theoretically relevant dimensions of institutional quality and stability. In line with Bang and Mitra (2011 and 2015) we find three dimensions of institutional *quality* relating to the levels of democracy, transparency, and policy credibility; whereas consistent with Jong-A-Pin (2009), we find dimensions to institutional *stability* relating to the levels of violence, protest, regime stability, and within-regime stability. In addition to these we find a dimension of institutional stability relating to *adverse regime change*.

Our measures of institutional quality and stability improve on existing measures in several important ways. First, by using data that are openly available, our results are easily reproducible. Also, by deriving weights on the observed variables based on an optimization algorithm, we produce factor variables that are relatively free of researcher-subjectivity bias. Next, since we use an orthogonal rotation of the loadings matrix, we derive institutional variables that are relatively uncorrelated with one another. In addition, by considering pairwise correlations between variables, and by using random forest imputation to make the most efficient use of available data, our measures are relatively robust to missing values. Finally, as we incorporate new data, our measures avoid the rigidness of pre-determine scoring weights, and can therefore adapt more flexibly to (in other words, "learn" from) new data.

3. Data and Methodology

3.1 Data

To measure the underlying latent variables that capture the various dimensions of institutional quality and stability, we consider 28 observable measures of institutional quality and stability from multiple sources over the time period of 1950-2014. First, to capture extreme and persistent forms of instability stemming from violent internal conflict, we include indicator variables for (a) ethnic and (b) revolutionary conflict from the Political Instability Task force (PITF).

Next, we include nine measures of governmental accountability from the DPI: (c) legislative party fractionalization; (d) government polarization, which measures the ideological distance between the legislative and executive branches of government; (e) executive tenure, measured as the number of years the current executive has been in office; (f) the number of changes in the number of veto players in the government; (g) government concentration, measured by the Herfindahl Index of the legislative governing coalition; (h) checks on power, measured by the number of veto points within the government; (i) the legislative and (j) executive indices of electoral competition; and (k) the presence of electoral fraud in the last election. To these, we add (l) the Polity 2 index of democracy and (m) regime durability from the Polity IV Project.

We also include 11 variables from the CNTS that capture other observable characteristics of institutional quality and stability: (n) the number of assassinations; (o) the number of labor strikes; (p) the number of major government crises; (q) the number of policital purges; (r) the number of riots; (s) the number of anti-government demonstrations; (t) the number of coups d'etat; (u) the number of major constitutional changes; (v) the number of major cabinet changes in the last five years; (w) the number of changes in the office of the effective chief executive in the last five years; and (x) the number of legislative elections in the last five years.

Finally, we add four indicators of economic and government crisis and policy instability from Reinhart and Rogoff (2011) to measure the effects of unpredictable changes in policies or economic risk. These types of crisis include (y) currency crises; (z) inflation crises; (aa) sovereign debt crises; and (bb) banking crises. We report the summary statistics for these 28 variables in Table 1. We include all 28 of these variables in a single factor analysis to detect patterns of common variance in their correlation matrix. We describe how factor analysis does this next.

3.2 Methodology

From the perspective of measuring the impact of institutional quality and stability on economic outcomes (such as growth), adding any combination of these variables may create a number of problems. Moreover, as noted by Acemoglu, et al. (2014), each of these variables is likely to be subject to considerable measurement error. However, in the case Acemoglu, et al. (2014) the main object is to pinpoint the exact *timing* of broad democratic transitions to examine the impact of democracy alone on economic growth. Furthermore, their analysis focuses only on whether a country is "democratic" or not. They construct a dichotomous index of democracy based on four main indices of democracy and political freedom common to the institutional literature.[1]

[1] Sources include: Polity; Freedom House; Cheibub, Gandhi, and Vreeland (2010); Boix, Miller, and Rosato (2012).

Then, the authors apply Papaioannou and Siourounis's (2006) method for locating the timing the episode of transitions.

Unfortunately, this method largely ignores the fact that various concepts of institutional quality from the various sources employed in the literature exhibit high degrees of collinearity. We could overcome the problem of multicollinearity performing principal component analysis on the variables, and interpreting the first principle component as "institutional quality." Alesina and Perotti (1996), Perotti (1996), and Keefer and Knack (1997) use this type of approach to investigate the influence of institutional quality on economic growth. In a similar vein, Alesina et al. (1996) constructs a one-dimensional index of institutional quality using logit regression. This approach would work well under the assumption that there is a single underlying concept defined broadly as "good institutions."

However, as argued by Jong-A-Pin (2009) in the context of political instability and by Bang and Mitra (2011) in the context of institutional quality, institutional quality and institutional stability are distinct concepts. Also, quality and stability may entail multiple dimensions in their own right. Moreover, the different dimensions of quality and stability may be highly correlated with one another. Thus, a one-dimensional index would fail to distinguish between various dimensions of institutional quality and stability. Additionally, given the scope for multicolinearity among these separate dimensions, measuring the impact of one dimension without controlling for the impact of the other is likely to subject any specification to considerable omitted variable bias.

3.2.1 Exploratory Factor Analysis

In light of these concerns, we perform exploratory factor analysis (EFA) on the institutional measures from the PITF, DPI, Polity, and CNTS. The EFA model is an unsupervised machine learning algorithm that expresses the vector of observed variables, $\boldsymbol{Y} = (y_1, ... y_K)$, as functions of a matrix of latent *common factors*, $\boldsymbol{F} = (f_1, \ldots, f_M)'$; a matrix of *factor loadings*, $\boldsymbol{\Lambda}$; a vector of *specific factors*, $\boldsymbol{\xi}$, and measurement error, $\boldsymbol{\varepsilon}$. The EFA model can be expressed as a vector of equations as follows:

$$(1) \qquad \begin{pmatrix} Y_1 \\ \vdots \\ Y_K \end{pmatrix} = \begin{pmatrix} \lambda_{11} f_1 \\ \vdots \\ \lambda_{1K} f_1 \end{pmatrix} + \cdots + \begin{pmatrix} \lambda_{M1} f_M \\ \vdots \\ \lambda_{MK} f_M \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_K \end{pmatrix}.$$ [2]

The EFA model then assumes that the errors are zero, and that the specific factors and measurement errors are independently and identically distributed.[3]

Notice that, given the number of restricting equations, EFA has the potential disadvantage of not obtaining a unique solution. While it is true that the solution to EFA is not unique, this is in fact an advantage, since all of the solutions are unique to a scaling constant. Additionally, since the EFA model focuses on covariance and not levels, we can express the model in terms of the vector of each observed variable's deviations from its respective mean, $\tilde{Y}_i = Y_i - \mu_i$. This allows us to focus only on the solutions that restrict the mean to be approximately zero and a variance to be approximately one.[4]

Imposing these restrictions, we can express equation (1) in terms of the covariance matrix:

$$(2) \qquad \boldsymbol{\Sigma}_{YY} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Xi}^2 + \boldsymbol{E}^2$$

or, if we normalize based on the product of the standard deviations of the variables in each respective pairwise covariance, we can express (1) in terms of the correlation matix:

$$(2') \qquad \mathbf{P}_{YY} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Xi}^2 + \boldsymbol{E}^2$$

---

[2] Contrast this with principle component analysis, where the components, $c_1,\ldots,c_k$ are functions of the observed variables, $y_1,\ldots,y_k$, and their coefficients.

[3] Expressed in technical terms, this assumes: (1) $E(\varepsilon) = 0$; (2) $E(\xi) = 0$; (3) $E(\xi\varepsilon) = 0$; (4) $E(f_m\xi) = 0$ for all factors, $m = 1, \ldots M$; (5) $E(f_m\varepsilon) = 0$ for all factors $m = 1, \ldots M$; (6) $E(\xi_{ik}\xi_{jl}) = 0$ for all observations $i \neq j$ and for variables, $k, l = 1, \ldots K$; and (7) $E(\varepsilon_{ik}\varepsilon_{jl}) = 0$ for all observations $i \neq j$ and variables $k, l = 1, \ldots K$.

[4] EFA does this by imposing the following additional restrictions: (1) $E(\boldsymbol{F}) = \mathbf{0}$; (R2) $V(\boldsymbol{F}) = \boldsymbol{I}$

where $\boldsymbol{\Xi}^2$ is a diagonal matrix comprised of unique variances for the observed variables and $\boldsymbol{E}^2$ is the diagonal matrix of variances of the measurement errors.

In this specification of the model, we see three distinct advantages to EFA. First, since it identifies common sources of covariance in the observed variables, it allows us to explore an underlying theoretical structure to the observed data. Secondly, and again because it focuses on the common sources of variation, the procedure allows us to purge measurement error from the observed variables constructing the latent factors. Third, these common sources of variation are, by construction, free of multicollinearity.[5]

3.2.2 Factor Extraction and Rotation

Next, we briefly exposit the EFA methodology to highlight how it achieves the goals of reducing the problem of multicollinearity and purging the observed institutional variables of measurement error. The discussion here is by no means complete, and focuses primarily on the extraction and rotation methods used for the results we report in Section 4. We refer interested readers to Chapter 10 of Rencher (1998) for more complete summary of the several different approaches to factor extraction and rotation, and to Gorusch (1983) for a comprehensive explanation of factor analysis methods.

3.2.3 Extraction

In obtaining the underlying latent factors, one faces the choice between several extraction and rotation methods. With respect to extraction the most prominent methods are principle axis, ordinary least squares (OLS), generalized least squares (GLS), and maximum likelihood (MLE). Of these, generalized least squares serves our purposes best because it (iteratively) gives more

---

[5] This result comes from the restriction that $V(\boldsymbol{F}) = \boldsymbol{I}$. However, in practice it is common to *rotate* the factors to obtain factors that, while not fully uncorrelated, lend themselves more readily to interpretation. We discuss factor rotation in the next subsection.

weight to observed variables that have higher levels of communality (and lower levels of uniqueness). The remainder of this subsection briefly explains the GLS extraction method.

The OLS factor extraction method starts with an initial value for the estimate of the errors, $\widehat{\boldsymbol{\Psi}}_0 = (\boldsymbol{\xi} + \boldsymbol{\varepsilon}) \cdot \boldsymbol{I}$. Then, it factors the diagonal elements of the correlation matrix of $\boldsymbol{Y}$ minus the errors, $\boldsymbol{H^2} = diag(\boldsymbol{P}_{YY}) - \widehat{\boldsymbol{\Psi}}_0$ to solve the equation:

(2") $\boldsymbol{H^2} \cong \boldsymbol{\Lambda\Lambda}'$.[6]

The solutions to (2") imply that $\boldsymbol{\Lambda\Lambda}'$ can be expressed as the spectral decomposition of $\boldsymbol{H}^2$, which means:

(3) $\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}' = \sqrt{\boldsymbol{\theta}_1}\boldsymbol{c}_1\boldsymbol{c}_1{}' + \cdots \sqrt{\boldsymbol{\theta}_m}\boldsymbol{c}_m\boldsymbol{c}_m{}'$,

where $\boldsymbol{c}_i$ is the i^{th} normalized eigenvector ($\boldsymbol{c}_1{}'\boldsymbol{c}_1 = 1$), and $\boldsymbol{\theta}_i$ is the corresponding eigenvalue of $\boldsymbol{H}^2$. OLS extraction then solves for the factor loadings that minimize the squared residuals, $\left(\boldsymbol{\Sigma}_{YY} - \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}'\right)^2$.

Generalized least squares extraction involves a similar process. The main way in which GLS extraction differs from OLS extraction is that the GLS method weights and re-weights $\boldsymbol{H}^2$ at each iteration based on the uniqueness. The result is that instead of solving the eigenvalues (or vectors) of the correlation matrix directly, we decompose the matrix, $\boldsymbol{\xi}^{-1}\boldsymbol{P}_{YY}\boldsymbol{\xi}^{-1}$. This leads to the desirable result that the model fits the highly common variables better than it does the highly unique ones (by giving terms in the correlation matrix involving highly unique variables lower weight).

3.2.3.1 Rotation

---

[6] For example, if we normalize the errors to be zero, then $\boldsymbol{H^2} \cong \boldsymbol{\Lambda\Lambda}' = \boldsymbol{\Sigma}_{YY}$, and the starting eigenvalues and eigenvectors for the first estimate of the loadings matrix come directly from the covariance (or correlation) matrix.

With respect to rotation, we have a choice between orthogonal methods, such as varimax, quartimax, equamax, varimin, and bifactor. Orthogonal methods have the advantage of maintaining orthogonality among the constructed factors, while oblique methods (such as promax, oblimin, simplimax, and biquartimin) have the advantage of imposing fewer restrictions on the model. For our purpose, varimax and oblimin perform well, but we have tested the robustness of our EFA by using alternative methods, and have obtained a nearly identical structure using quartimax, promax, and quartimin.[7] The remainder of this subsection briefly describes the varimax family rotation method.

The concept of factor rotation in EFA exploits the fact that the solution to the EFA problem is not unique, except to a scaling constant. Therefore, the population loading matrix $\boldsymbol{\Lambda}$ can be rotated using any orthogonal matrix $\boldsymbol{T}$ to $\boldsymbol{\Lambda}^{*} = \boldsymbol{\Lambda T}$ without violating any of the assumptions of the model.[8] As its name suggests, the varimax method maximizes the variances of the factor loadings, $V = \sum_{i=1}^{n}\left(\lambda_{i,k} - \bar{\lambda}_{i,k}\right)^2$ for each factor k (Kaiser, 1958).

4. Measuring Institutions

4.1 A Synthesis of Institutional Quality and Stability

In our initial exploration of the data, we seek to measure the underlying institutional factors of all 28 variables from the PITF, DPI, Polity, CNTS, and Reinhart and Rogoff. We report the results of this EFA (using the GLS extraction method and varimax rotation) in Table 2. In the table, variables with factor loadings close to zero for a particular factor contribute very little to the

[7] We will make these results available on request.

[8] Notice that since $\boldsymbol{T}$ is orthogonal, it has the property that $\boldsymbol{TT'} = \boldsymbol{I}$, and so the matrix $\widehat{\boldsymbol{\Lambda}}\boldsymbol{T}$ is also a valid solution to equation (2").

construction of that measure, and so we have omitted factor loadings with an absolute magnitude less than 0.1 from the results. We also report the eigenvalue and the proportion of the total variance explained by each factor at the bottom of the table.

Nine factors emerge from this analysis: (1) democracy; (2) protest; (3) within-regime instability; (4) credibility; (5) adverse regime change; (6) institutional flexibility; (7) regime instability; (8) violence; and (9) transparency. We briefly explain the interpretations of the factors below.

(1) Democracy measures the extent to which political leaders are chosen by the people and to which the political process incorporates the preferences of a diverse set of interests. Variables that comprise this dimension of institutional quality include (with rotated factor loadings in parentheses): the executive and legislative indices of electoral competition (0.971 and 0.874, respectively); the Polity index (0.784); checks on power (0.735); legislative fractionalization (0.657); government polarization (0.558); and executive tenure (-0.354). Higher scores of this factor indicate more democratic institutions.

(2) **Protest** is constructed primarily from the numbers of riots (0.838), demonstrations (0.725), and strikes (0.383) in society. Higher numbers indicate greater unrest.

(3) **Within**-regime instability incorporates factors that may indicate sources of instability stemming from fissures in ideology within the regime. Hence, it is based on legislative concentration as measured by the Herfindahl index (-0.783), legislative fractionalization (0.680) and political polarization (0.448) indices from the DPI. Countries with more fractious, and therefore presumably less stable, governments receive higher values.

(4) **Credibility** relates to the confidence the public is able to place in the predictability of government policies and the relative freedom from expropriation of profits and property by the

government. Inflation crises (0.715) and currency crises (0.686) combine to define this factor. Higher scores for this factor correlate with *lower* levels of predictability and credibility.

(5) **Adverse** regime change captures the dimensions of regime instability that we associate with sudden, and sometimes violent changes in the regime. As such, the number of coups (0.537), changes in the executive (0.532), cabinet changes (0.437), constitutional changes (0.418), and political purges (0.299) variables combine to construct this variable. Higher scores on this factor correlate with greater *instability*.

(6) **Flexibility** of institutions relates to the institutions' ability to adapt and accommodate change. Institutions with higher overall rates of peaceful turnover will tend to be more flexible. Hence, this factor captures flexibility through executive turnover (-0.672), changes in veto players (0.417), and the level of democracy as measured by the Polity2 index (-0.316). Higher scores of this factor indicate more flexible institutions.

(7) **Regime** instability corresponds to relatively peaceful changes in power in the regime. The number of executive changes (0.546) and major cabinet changes (0.496), and legislative elections compose this index. Higher values reflect greater instability. Whereas the *adverse regime change* factor captured the components of these turnover variables that correlate with violent events such as coups and purges, this variable extracts the variance in these variables in common with electoral changes in government.

(8) **Violence** captures the risk for internal conflict. The prevalence of revolutionary civil war (0.528), assassinations (-0.348), and ethnic conflict (0.330) comprise this factor. Higher scores indicate greater instability.

(9) **Transparency** is the degree to which the regime is able to curb corruption and fraud, and to efficiently execute bureaucratic functions. Governments that fail to do so will not

endure. Thus, in our model, the variation that regime durability (0.440) and government polarization (0.446) share with electoral fraud (-0.331), checks on power (0.295), and the Polity index (0.226) comprise this factor.

Figure 1 offers a visualization of these relationships in the form of a factor diagram. In this diagram, we have constrained each observed variable to connect to at most one of the factors – the one with the highest loading for that variable. We have also constrained the diagram to only include connections if the associated factor loadings exceed 0.30. Connections in red correspond to negative associations.

4.2 Enhancements: Random Forest Imputation

One problem with existing institutional datasets is that the data are not available for a very long time series. By contrast, the effects of institutions and institutional change might only occur over the course of the long run. Conveniently, some computational algorithms for factor analysis include the option for imputing missing values using the median or the mean. While this is attractive, a better solution to the problem of missing values (and short time horizons) would be to make full use of the distribution of the variables with missing values conditional on observed values for all of the other variables. This is the essence of what random forest imputation seeks to do.

Random forests is a supervised machine learning method for predict (or classify for discrete outcomes) the value of an outcome based on a set of predictors using an ensemble of classification and regression trees. Random forest imputation imputes values based on a random forest prediction algorithm. In this section, we briefly describe how the random forest algorithm imputes values for missing observations, and present the results of the factor analysis using the imputed values.

In contrast to other methods, in which the starting point of the analysis is to estimate the conditional distribution of the missing-valued variables, random forests (and other tree-based algorithms that impute missing values) begin from trying to predict the conditional distribution of an outcome variable, $\boldsymbol{Y}$, (which does not have missing values) based on a set of predictors, $\boldsymbol{X}$. In our case, we use a weighted average of conflict, protest variables, and coups constitutional changes, and fraud as the outcome variable, since these variables contain very few missing values.[9] The predictors are all of the institutional variables from our factor analysis.

The random forest algorithm builds many – in our case 500 – regression trees to predict the outcome variable based on the predictors. To build each tree, the algorithm first considers the entire sample, and randomly selects a subset (usually of size 2 to 4) of the predictors.[10] Out of that subset of predictors, the algorithm considers all of the possible discrete splits of those variables and selects the variable and the cutoff for that variables in such a way as to minimize the mean square error (MSE) (or the Gini Index for discrete outcomes) of the outcome variable. This creates two subsets of the data, which the algorithm seeks to split successively further using newly-selected variables (with replacement) until the tree meets a terminating condition (either a minimum improvement in MSE or maximum tree size).

At each node, if the algorithm encounters missing values for any predictor, it responds by substituting the conditional *median* (or mode in the case of a discrete predictor) for that subnode, which is conditional on all of the information prior to that subnode, including both the other

[9] For the small number of observations that the fraud, protest, coups, and constitutional change variables do return missing values we substitute a value of zero for the missing values for the purposes of constructing our outcome. Then, take the first principle component of these variables as our outcome in the prediction algorithm.

[10] In a sense, this method is similar to "bootstrapping" (also known as "bagging"). However in contrast to the "tree bagging" approach, which randomly selects *observations* from the sample, random forests randomly select predictors, or features, that predict the outcome. Some have referred to this part of the random forest algorithm as "feature bagging."

predictors and the outcomes. Finally, after constructing the entire forest, the algorithm gives each tree a "vote" over which value to use in place of the missing values for each observation. We report the summary statistics for the dataset with the imputed values added in Table 3.[11]

Using the random forest imputed values for the entire sample period of 1950-2014, we have once again performed the 28-variable EFAs and report the results in Table 4, and offer the factor diagram in Figure 2. We see that, with the exception of minor differences in the specific values for the factor loadings, and slight differences in the order of the factors, the EFA yields the same nine factors as we have described in Section 4.1.

4.3 Discussion

One useful feature of our dataset is that the variables in it bear a striking similarity to the four political instability factors (violence, protest, regime, and within) obtained by Jong-a-Pin (2009) and three institutional quality factors (democracy, transparency, and credibility) obtained by Bang and Mitra (2011). For this reason, we have applied the same terms in our interpretation of these factors. Of the remaining two, adverse regime change (the variation that measures of regime turnover share with coups and purges) fits well with the definition of regime change laid out by the PITF; and institutional flexibility coincides well with the definition of institutional flexibility given by Davis (2010). In this sense, our results are quite consistent with previous contributions to the literature on institutions.

We have also taken care to ensure that our dataset could be reproduced, modified, or extended by any researcher. We almost exclusively use observed variables that come from datasets

[11] One useful property of the random forest imputation method is that the imputed values lie within the range of the existing sample. Parametric methods (such as multiple imputation) may impute values that lie far outside of the range of the existing sample.

that are freely available online: the PITF conflict variables and the Polity IV variables are freely available from the Center for Systematic Peace (http://www.systemicpeace.org/inscrdata.html); the Database of Political Institutions, formerly available from the World Bank, is now available from the Inter-American Development Bank (https://publications.iadb.org/handle/11319/7408); and the Reinhart and Rogoff crisis data can be downloaded from Carmen Reinhart's webpage (http://www.carmenreinhart.com/data/) and is well-documented in the accompanying paper (Reinhart and Rogoff, 2011). The fully-updated CNTS data does come at a nominal fee,[12] but there is also a trial version that covers through 2008 available at the databanks website (https://www.databanksinternational.com/Trial/). However, even though we use the paid version of the data, the variables on political structure and social unrest that we use *could* be closely reproduced by hand from constitutional documents and news reports (albeit with substantial time and effort). Another step we take to improve the reproducibility of our factors is that we not only make the final sets of factor scores freely available for download, but we also publish the R source code that we use to generate the factors. The dataset and source code that generates the dataset are available at: https://www.researchgate.net/project/Machine-Learning-Political-Indicators.

Another important feature of our dataset is that the factor analysis allows the loading weights to adapt as we add updated information to the dataset. This means that as new years of data become available for the observed variables, the factor loadings (and therefore the factor scores) will change, both for the newly observed years and for years going back. This might seem like a drawback: Why shouldn't the democracy score for Brazil in 1984 remain constant as I add new years of data? The answer lies in the premise from which our dataset started – that the concepts

[12] The current price for academic users is $550 for a single license; fees for non-academic users are significantly higher.

of institutional quality and stability that we wish to observe are latent variables, and thus hidden from direct observation. Any approach to measuring such concepts of institutional quality and stability should concede that any attempts to measure these concepts are therefore only guesses – priors, to use Bayesian terms – of the true values. From this (Bayesian) perspective, it only makes sense that we should update our judgment of Brazil's past institutions when new information about the correlation between the various observed characteristics comes to light. Our method allows our *approach* to remain consistent, while allowing our *priors* (the factor scores) to perpetually update.

We anticipate the possibility that some potential users of our dataset might approach the introduction of yet another database with some skepticism. To address this skepticism, we compare the scores from our analysis with one of the more common databases, the International Country Risk Guide (ICRG). The ICRG Political Risk table defines 12 measures of institutions, nine of which (democratic accountability, government stability, investment profile, corruption, bureaucratic quality, socioeconomic conditions, military in politics, religion in politics, and rule of law) measure institutional quality; and three of which (internal conflict, external conflict, and ethnic tensions) measure instability, namely various forms of conflict. We report the correlation coefficients between our factor scores and the ICRG Political Risk Rating components in Table 5.

Our measures of institutional quality correlate closely with the dimensions of institutional quality from the ICRG that would be consistent with our interpretations of them. *Democracy* correlates with "democratic accountability", and somewhat with "military in politics"; *credibility* correlates with "investment profile,"[13] "socioeconomic conditions,"[14] "government stability,"[15]

[13] The Political Risk Services (PRS) Group defines investment profile as "an assessment of factors affecting the risk to investment." It includes contract viability/expropriation, profits repatriation, and payment delays.
[14] "[A]n assessment of the socioeconomic pressures…that could constrain government action or fuel…dissatisfaction."
[15] "[A]n assessment both of the government's ability to carry out its declared program(s), and…stay in office."

and "bureaucratic quality"[16]; *transparency* correlates with many of the measures, but is most closely associates with "bureaucratic quality," "corruption,"[17] "rule of law," [18] and "democratic accountability"; and *flexibility* correlates somewhat with "democratic accountability," but also correlates with other measures not captured by the ICRG.

Numerous studies use the ICRG variables – especially the investment profile index as a measure of property rights – in measuring the impact of institutions on economic development. Our measures – and especially our "credibility" factor – improves on this approach by being a cleaner measure of policy stability in the sense that it is relatively uncorrelated with other concepts of institutional quality, such as democracy or transparency. Tables 6 and 7 report the correlations among the ICRG variables and our factor scores, respectively, and supports this claim. Out of the 66 pairwise correlation coefficients among the ICRG variables in Table 6, 48 (73%) correlate with a coefficient above 0.3; and 21 correlate with a coefficient above 0.5. By contrast, out of 45 pairwise correlation coefficients among our factor scores in Table 7, only 1 correlates with a coefficient above 0.3; none correlates with a coefficient above 0.5.

5. Conclusion

In this paper we find that the theoretical concepts of institutional quality and stability discussed in the literature can be described by seven common factors that we derive using an EFA of 30 observable measures of institutional character. We also find that most aspects of institutional quality can be captured very efficiently using plainly observable aspects of a country's

[16] Bureaucratic quality is highest "where the bureaucracy has the strength and expertise to govern without drastic changes in policy or interruptions in government services."
[17] While the target of this variable is self-explanatory, we should note that higher scores move towards *less* corruption.
[18] Law measures "the strength and impartiality of the legal system are considered;" order assesses "popular observance of the law."

constitutional framework, and measurable outcomes of the political process. This gives future researchers the benefit of a reproducible methodology for constructing the underlying concepts that define “good institutions.”

One important policy implication of this research is that “good institutions” involves many things. Concepts of quality such as democracy, transparency, and credibility, as well as various forms of stability such as violence, protest, regime stability, and measures of fractious-ness within the regime all contribute to the general idea of “goodness” of institutions. Moreover, we have shown that many of the observed characteristics that previous studies have used to proxy for these measures are highly correlated, and hence it is difficult to disentangle the effects of one concept from the effects of the other concepts without employing some sort of signal extraction technique, as we have done using EFA.

More importantly, our analysis has shown that, for any given concept of “goodness” of institutions, there may be several structural characteristics that define it. Take democracy, for example. While elections are an important part of adopting “good” democratic institutions, so too are other other characteristics, such as implementing checks and balances on the power of one branch of government vis-à-vis the others; allowing and electing a more pluralistic, and therefore less concentrated legislature; administering an efficient and transparent bureaucracy; and even having structures that support the division of power among polarized ideologies between the legislature and the chief executive correlates highly with “good democracy.”

Future avenues for research in this area include more carefully analyzing the data to search for *de jure* institutional characteristics that correlate with property rights in order to recover the concept of credibility from the data. Also, future research could use some of the alternate methods for constructing and imputing concepts of institutional quality to analyze other economic outcomes, such as economic growth. This would allow us to compare the different construction methods and better understand which concepts (and therefore which structural characteristics) are most important for achieving desirable societal outcomes.

References

Adsera, A., & Boix, C. (2002). Trade, democracy, and the size of the public sector: the political underpinnings of openness. International Organization, 56(02), 229-262.

Aisen, A., & Veiga, F. J. (2013). How does political instability affect economic growth?. European Journal of Political Economy, 29, 151-167.

Alesina, A., & Perotti, R. (1996). Income distribution, political instability, and investment. European economic review, 40(6), 1203-1228.

Asiedu, E. (2006). Foreign direct investment in Africa: The role of natural resources, market size, government policy, institutions and political instability. World Economy, 29(1), 63-77.

Bang, J., T. and Mitra, A. (2011). Brain drain and the institutions of governance: educational attainment of immigrants to the US, 1988-1998. Economic Systems, 35(3): 335-54.

Bang, J., T. and Mitra, A. (2013). Civil war, ethnicity, and the migration of skilled labor. Eastern Economic Journal, 39(3): 387-401.

Basu, A., Bang, J. T., and Shughart II, W. F. (2015). Predicting state failure: a machine learning approach. Mimeo, St. Ambrose University.

Busse, M., & Hefeker, C. (2007). Political risk, institutions and foreign direct investment. European journal of political economy, 23(2), 397-415.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic natural disasters and economic growth. Review of Economics and Statistics, 95(5), 1549-1561.

Coase, R. H. (1960). The problem of social cost. Journal of Law and Economics, 3(1), 1-44.

Davis, L. (2010). Institutional flexibility and economic growth. Journal of Comparative Economics 38(3), 306-320.

Demsetz, H. (1967). Toward a theory of property rights. American economic review, 347-359.

Easterly, W. (2001). Can Institutions Resolve Ethnic Conflict? Economic Development and Cultural Change, 49(4), 687-706.

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. Quarterly Journal of Economics, 1203-1250.

Esfahani, H. S., & Ramírez, M. T. (2003). Institutions, infrastructure, and economic growth. Journal of Development Economics, 70(2), 443-477.

Fatás, A., & Mihov, I. (2013). Policy volatility, institutions, and economic growth. Review of Economics and Statistics, 95(2), 362-376.

Ghate, C., Le, Q. V., & Zak, P. J. (2003). Optimal fiscal policy in an economy facing sociopolitical instability. Review of Development Economics, 7(4), 583-598.

Gorsuch, R. L. (1983). Factor Analysis, 2nd. Hillsdale, NJ: LEA.

Grogan, L., & Moers, L. (2001). Growth empirics with institutional measures for transition countries. Economic systems, 25(4), 323-344.

Jong-A-Pin, R. (2009). On the measurement of political instability and its impact on economic growth. European Journal of Political Economy, 25(1), 15-29.

Kahn, M. E. (2005). The death toll from natural disasters: the role of income, geography, and institutions. Review of economics and statistics, 87(2), 271-284.
Keefer, P., & Knack, S. (2002). Polarization, politics and property rights: Links between inequality and growth. Public Choice, 111(1-2), 127-154.
Knack, S., & Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. Economics & Politics, 7(3), 207-227.
Langbein, L., & Knack, S. (2010). The worldwide governance indicators: six, one, or none? Journal of Development Studies, 46(2), 350-370.
Leblang, D. A. (1996). Property rights, democracy and economic growth. Political Research Quarterly, 49(1), 5-26.
Li, Q., & Resnick, A. (2003). Reversal of fortunes: Democratic institutions and foreign direct investment inflows to developing countries. International Organization, 57(01), 175-211.
Lin, J. Y., & Nugent, J. B. (1995). Institutions and economic development. Handbook of Development Economics, 3, 2301-2370.
Milner, H. V., & Kubota, K. (2005). Why the move to free trade? Democracy and trade policy in the developing countries. International Organization, 59(01), 107-143.
Mitra, A., and Bang, J. T. (2015). Post-conflict institutions and the incidence of conflict: a cross-national analysis. Levy Institute Working Paper.
Mitra, A., Bang, J. T. and Wunnava, P. (2014). Financial liberalization and the selection of emigrants: a cross-national analysis. Empirical Economics, 47(1): 199-226.
Morrow, J. D., Siverson, R. M., & Tabares, T. E. (1998). The political determinants of international trade: the major powers, 1907–1990. American Political Science Review, 92(03), 649-661.
North, D. C. (1981). Structure and Change in Economic History. New York: Norton.
North, D. C. (1989). Institutions and economic growth: An historical introduction. World Development, 17(9), 1319-1332.
North, D. C., & Weingast, B. R. (1989). Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. Journal of economic history, 49(04), 803-832.
Olson, M. (1965). The Logic of Collective Action: Public Goods and the Theory of Group Cambridge: Harvard University Press.
Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press.
Perotti, R. (1996). Growth, income distribution, and democracy: what the data say. Journal of Economic Growth, 1(2), 149-187.
Przeworski, A., & Limongi, F. (1993). Political regimes and economic growth. Journal of Economic Perspectives, 51-69.
Reinhart, C. and K.S. Rogoff (2011). From financial crash to debt crisis. American Economic Review, 101(5), 1676-1706.
Rencher, A. C. (1998). Multivariate statistical inference and applications. New York: Wiley.
Rodrik, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. Journal of economic growth, 4(4), 385-412.
Sala-i-Martin, X. X. (1997). I just ran two million regressions. American Economic Review, 87(2), 178-183.
Weingast, B. R. (1993). Constitutions as governance structures: The political foundations of secure markets. Journal of Institutional and Theoretical Economics, 286-311.

Weingast, B. R. (1995). The economic role of political institutions: market-preserving federalism and economic development. Journal of Law, Economics, & Organization, 1-31.
Williamson, O. E. (1979). Transaction-cost economics: the governance of contractual relations. Journal of Law and Economics, 233-261.

Tables

Table 1: Summary Statistics of Observed Sample

Statistic	Source	N	Mean	St. Dev.	Min	Max
Ethnic Conflict	PITF	10,577	0.081	0.273	0	1
Revolutionary Conflict	PITF	10,577	0.051	0.220	0	1
Legislative Fractionalization	DPI	5,294	0.472	0.302	0	1
Government Polarization	DPI	5,640	0.359	0.730	0	2
Executive Tenure	DPI	6,232	7.704	8.056	0	47
Change in Veto Players	DPI	5,934	0.113	0.273	0	1
Government Concentration	DPI	5,360	0.813	0.267	0.02	1.00
Checks on Power	DPI	6,072	2.527	1.699	1	18
Legislative Electoral Competition	DPI	6,231	5.455	2.115	1	7
Executive Electoral Competition	DPI	6,226	5.213	2.154	1	7
Electoral Fraud	DPI	5,203	0.127	0.333	0	1
Polity2 Democracy Index	Polity	8,723	0.586	7.483	-10	10
Regime Durability	Polity	8,801	21.700	27.974	0	204
Assassinations	CNTS	9,950	0.166	0.875	0	26
General Strikes	CNTS	9,950	0.114	0.533	0	13
Government Crises	CNTS	9,950	0.163	0.502	0	7
Political Purges	CNTS	9,950	0.122	0.632	0	34
Riots	CNTS	9,950	0.463	1.845	0	55
Anti-Government Demonstrations	CNTS	9,950	0.596	2.417	0	74
Coups d'Etat	CNTS	9,891	0.026	0.165	0	2
Constitutional Changes	CNTS	9,874	0.084	0.287	0	4
Number of Cabinet Changes	CNTS	9,753	0.428	0.595	0	7
Changes in Executive	CNTS	9,887	0.189	0.456	0	7
Legislative Elections	CNTS	9,843	0.221	0.418	0	3
Currency Crises	RR	4,062	0.175	0.391	0	2
Inflation Crises	RR	4,067	0.151	0.358	0	1
Domestic Sovereign Debt Crises	RR	4,067	0.028	0.164	0	1
Banking Crises	RR	4,067	0.111	0.314	0	1

Table 2: Rotated Factor Loadings, Observed Sample

	Democracy	Protest	Within	Credibility	Adverse	Flexibility	Regime	Violence	Transparency
Ethnic Conflict								0.330	
Revolutionary Conflict								0.528	
Leg. Fractionalization	0.657		0.680			0.182	0.120		-0.151
Gov. Polarization	0.458		0.448						0.446
Executive Tenure	-0.309					-0.672		-0.103	
Change in Veto Players						0.417			
Gov. Concentration	-0.207		-0.783						
Checks on Power	0.735		0.235						0.295
Leg. Electoral Comp.	0.874		0.111		-0.129		0.112		-0.135
Ex. Electoral Comp.	0.971								
Electoral Fraud						-0.256		0.192	-0.331
Polity2 Index	0.784		0.151			0.316	0.152		0.226
Regime Durability	0.133			-0.161	-0.204			-0.112	0.440
Assassinations		0.125						0.348	
General Strikes		0.383		0.200					
Government Crises		0.104						0.164	
Political Purges		0.191			0.299		0.150	0.126	
Riots		0.838						0.100	
Anti-Gov. Demonstr.		0.725						0.134	
Coups d'Etat					0.537				
Constitutional Changes	-0.115				0.418				
Cabinet Changes					0.437		0.496		
Changes in Executive					0.532	0.121	0.546		
Legislative Elections							0.356		
Currency Crises				0.686					
Inflation Crises	-0.123			0.715				0.108	-0.110
Sovereign Debt Crises				0.225			0.109	0.208	
Banking Crises				0.220			0.113	0.130	
Eigen Value	4.861	2.717	1.820	1.727	1.318	1.301	1.151	1.107	1.020
Proportion of Variance	0.133	0.053	0.05	0.043	0.042	0.032	0.028	0.027	0.027

Table 3: Summary Statistics of Imputed Sample

Statistic	Source	N	Mean	St. Dev.	Min	Max
Ethnic Conflict	PITF					
Revolutionary Conflict	PITF					
Legislative Fractionalization	DPI					
Government Polarization	DPI					
Executive Tenure	DPI					
Change in Veto Players	DPI					
Government Concentration	DPI					
Checks on Power	DPI					
Legislative Electoral Competition	DPI					
Executive Electoral Competition	DPI					
Electoral Fraud	DPI					
Polity2 Democracy Index	Polity					
Regime Durability	Polity					
Assassinations	CNTS					
General Strikes	CNTS					
Government Crises	CNTS					
Political Purges	CNTS					
Riots	CNTS					
Anti-Government Demonstrations	CNTS					
Coups d'Etat	CNTS					
Constitutional Changes	CNTS					
Number of Cabinet Changes	CNTS					
Changes in Executive	CNTS					
Legislative Elections	CNTS					
Currency Crises	RR					
Inflation Crises	RR					
Sovereign Debt Crises	RR					
Banking Crises	RR					

Table 4: Rotated Factor Loadings, Imputed Sample

	Democracy	Protest	Credibility	Regime	Within	Adverse	Violence	Flexibility	Transparency
Ethnic Conflict							0.299		
Revolutionary Conflict							0.672		
Leg. Fractionalization	0.72				0.499			0.146	
Gov. Polarization	0.56				0.383				0.36
Executive Tenure	-0.356							-0.648	
Change in Veto Players								0.489	
Gov. Concentration	-0.296				-0.872				
Checks on Power	0.786				0.183			0.117	0.208
Leg. Electoral Comp.	0.892								
Ex. Electoral Comp.	0.955								
Electoral Fraud							0.231	-0.139	-0.291
Polity2 Index	0.765			0.116				0.233	0.213
Regime Durability	0.161		-0.162				-0.102	-0.123	0.54
Assassinations		0.148					0.325		
General Strikes		0.399	0.132						
Government Crises		0.116					0.145		
Political Purges		0.217		0.277		0.137	0.134	0.116	
Riots		0.826							
Anti-Gov. Demonstr.		0.725					0.114		
Coups d'Etat				0.173		0.876			
Constitutional Changes	-0.168			0.213		0.306	0.104		-0.131
Cabinet Changes				0.72		0.124			-0.128
Changes in Executive				0.653		0.262		0.1	
Legislative Elections				0.375					
Currency Crises			0.733						
Inflation Crises	-0.163		0.763				0.102		-0.123
Sovereign Debt Crises			0.235				0.226		
Banking Crises									
Eigen Value	5.043	2.684	1.894	1.799	1.370	1.273	1.127	1.058	1.018
Proportion of Variance	0.146	0.054	0.05	0.046	0.044	0.036	0.032	0.031	0.025

Table 5: Correlation of Factor Scores with ICRG Variables

	Democracy	Protest	Within	Credibility	Adverse	Flexibility	Regime	Violence	Transparency
Dem. Acct.	0.501	-0.133	0.233	-0.249	-0.151	0.377	0.132	-0.197	0.509
Gov. Stab.	0.032	-0.125	0.139	-0.310	-0.169	0.007	-0.036	-0.265	0.121
Inv. Prof.	0.209	-0.139	0.149	-0.414	-0.140	0.162	0.039	-0.324	0.319
Corrupt	0.229	-0.063	0.117	-0.167	-0.107	0.166	0.045	-0.244	0.511
Bur. Qual	0.259	-0.015	0.152	-0.308	-0.100	0.122	0.086	-0.289	0.538
Int. Conf.	0.129	-0.185	0.153	-0.288	-0.078	0.112	0.064	-0.601	0.372
Ext. Conf.	0.212	-0.134	0.232	-0.136	0.014	0.095	0.030	-0.263	0.306
Eth. Ten.	0.195	-0.177	0.070	-0.053	-0.070	0.169	0.084	-0.364	0.144
Soc. Econ.	0.108	-0.026	0.079	-0.354	-0.111	0.103	0.034	-0.300	0.457
Mil. Pol.	0.305	-0.133	0.123	-0.270	-0.092	0.196	0.080	-0.407	0.491
Rel. Pol.	0.226	-0.141	0.006	-0.080	-0.064	0.256	0.054	-0.345	0.241
Rule of Law	0.153	-0.148	0.217	-0.294	-0.144	0.091	0.102	-0.436	0.510

Table 6: Correlations among ICRG Variables

	Dem. Acct.	Gov. Stab.	Inv. Prof.	Corrupt	Bur. Qual	Int. Conf.	Ext. Conf.	Eth. Ten.	Soc. Econ.	Mil. Pol.	Rel. Pol.
Gov. Stab.	0.200										
Inv. Prof.	0.485	0.554									
Corrupt	0.564	0.134	0.237								
Bur. Qual	0.645	0.297	0.479	0.703							
Int. Conf.	0.491	0.379	0.430	0.487	0.565						
Ext. Conf.	0.356	0.273	0.278	0.329	0.338	0.591					
Eth. Ten.	0.258	0.187	0.180	0.305	0.276	0.557	0.309				
Soc. Econ.	0.446	0.287	0.621	0.521	0.659	0.470	0.258	0.245			
Mil. Pol.	0.685	0.228	0.471	0.653	0.754	0.650	0.403	0.385	0.566		
Rel. Pol.	0.294	0.115	0.235	0.361	0.323	0.429	0.287	0.341	0.326	0.390	
Rule of Law	0.569	0.357	0.444	0.698	0.735	0.745	0.467	0.439	0.624	0.713	0.355

Table 2: Correlations among Institutional Factor Scores

	Democracy	Protest	Within	Credibility	Adverse	Flexibility	Regime	Violence
Protest	-0.142							
Within	0.107	-0.086						
Credibility	0.069	0.016	-0.070					
Adverse	0.092	0.010	0.012	0.149				
Flexibility	0.332	-0.007	-0.086	-0.035	-0.018			
Regime	-0.050	0.016	-0.004	-0.086	0.236	0.068		
Violence	0.020	0.108	-0.099	0.173	0.035	-0.075	-0.033	
Transparency	0.224	-0.095	0.261	-0.132	0.047	0.083	-0.005	-0.136

Figures

Figure 1: Factor Diagram, Raw Data

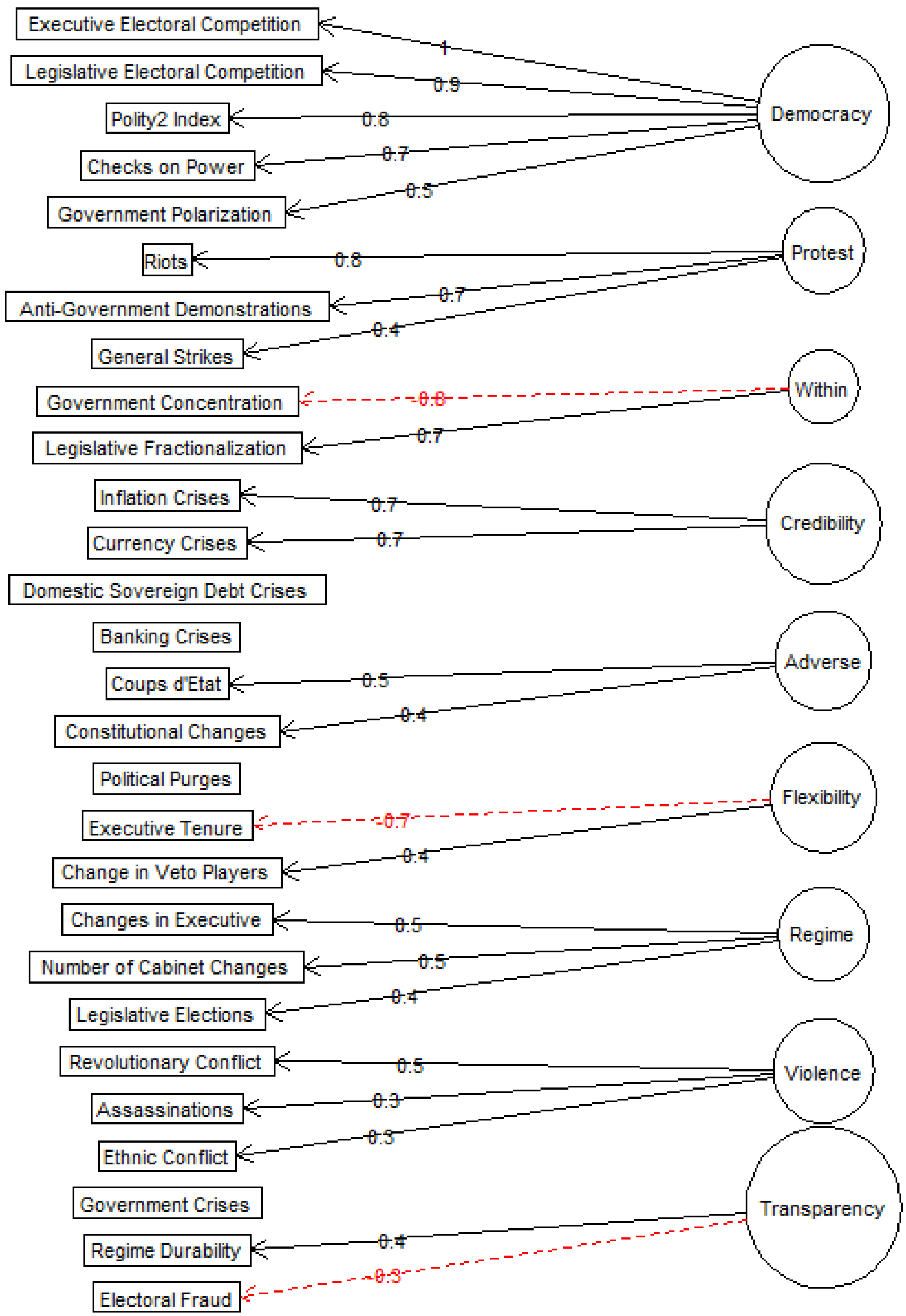

Figure 2: Factor Diagram, Imputed Data