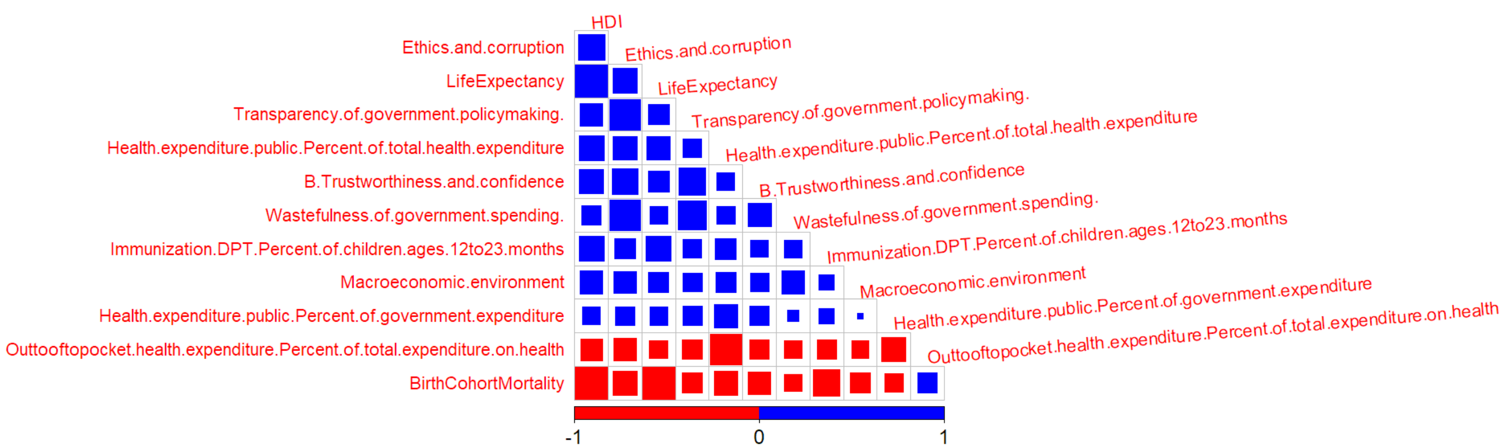
LEARNING ABOUT THE RISK OF BUDGET UNCERTAINTY

NICHOLAS J. MORRIS

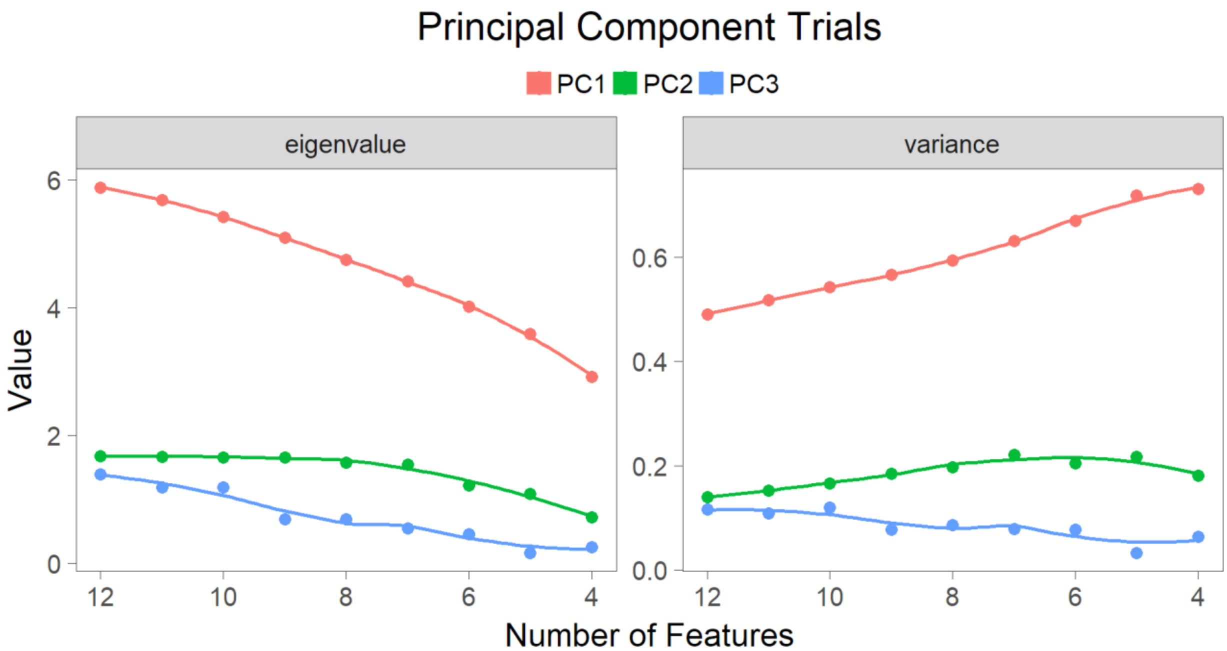
Rochester Institute of Technology NY, USA

METHODOLOGY

The correlation between macro-economic features of interest is shown below. This data represents 147 countries between the years of 2006 and 2016. This represents most of the 194 countries in our target market. Not all countries and years have values for every feature, representing unreported data. The correlation matrix below shows severe multi-collinearity which is promising since we want to represent multiple dimensions with a single axis.



An experiment was conducted to ensure that the first principal component capture a sufficiently large majority of the explained variance in the data. In each trial of the experiment, principal components analysis was run on the data to determine which feature of the data set whose variance was explained least by the first principal component. Finding this feature can be determined by looking at the first principal component loadings of each feature and finding the loading that is smallest in magnitude. The square of the principal component loadings represents the percent of a feature's total variance explained by the principal component. The figure below shows how the eigenvalue and explained variance of the first 3 principal components change as features were removed from the data set for having the smallest first principal component loading.



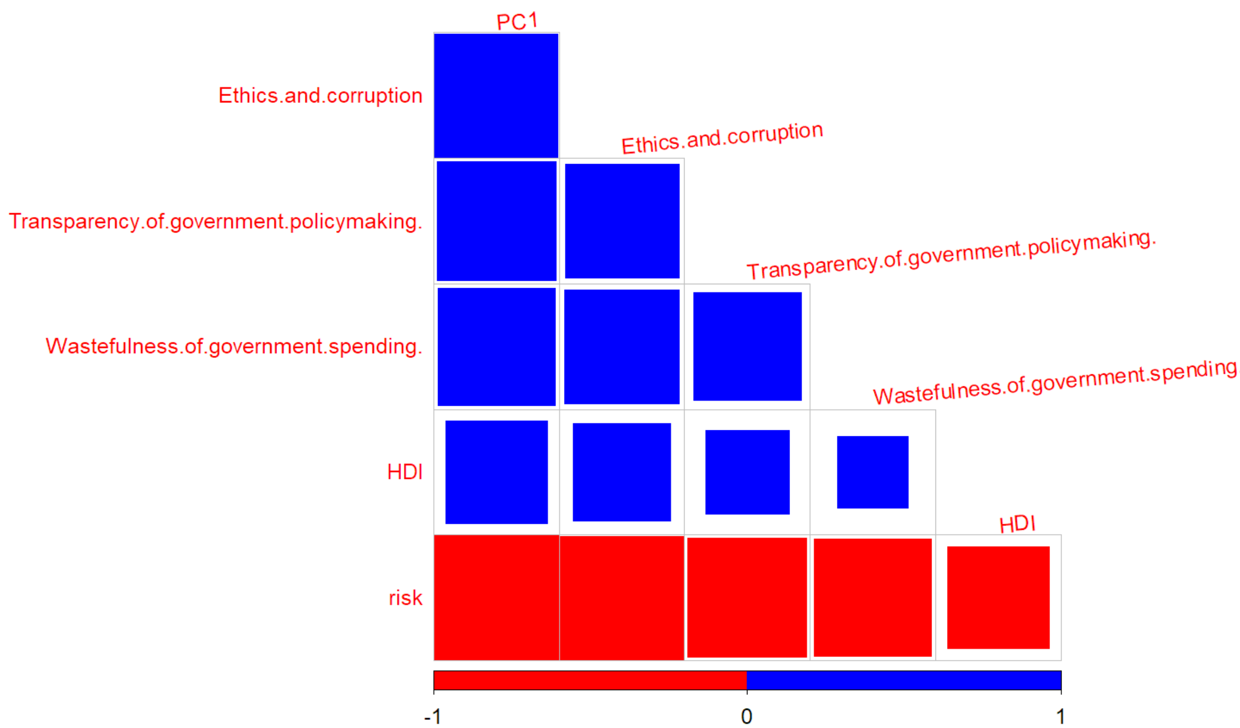
The table below details which feature was removed during each trial. In every trial, except the last one, the feature with the minimum PC1 loading from the previous trial is removed. The last trial did not remove the feature with the minimum PC1 loading because the data set would have consisted of Life Expectancy, Birth Mortality, HDI, and Ethics and Corruptions. HDI is the Human Development Index which is composed of life expectancy and infant mortality rates. So, Life Expectancy and Birth Mortality were removed to give previous features another opportunity to improve the explained variance of the first principal component.

Features	Feature with Minimum PC1 Loading	PC1 Loading	Action
12	Health expenditure public Percent of government expenditure	-0.1955349	Run all 12 Features
11	Macroeconomic environment	-0.2331182	Remove Health expenditure public Percent of government expenditure
10	Out of pocket health expenditure Percent of total expenditure on health	0.2651379	Remove Macroeconomic environment
9	Immunization DPT Percent of children ages 12 to 23 months	-0.2790886	Remove Out of pocket health expenditure Percent of total expenditure on health
8	Health expenditure public Percent of total health expenditure	0.2864807	Remove Immunization DPT Percent of children ages 12 to 23 months
7	Wastefulness of government spending	0.3284764	Remove Health expenditure public Percent of total health expenditure
6	Trustworthiness and confidence	0.3512945	Remove Wastefulness of government spending
5	Transparency of government policy making	0.3535485	Remove Trustworthiness and confidence
4	HDI	0.3853966	Remove Life Expectancy, Remove Birth Mortality, Add Wastefulness of government spending

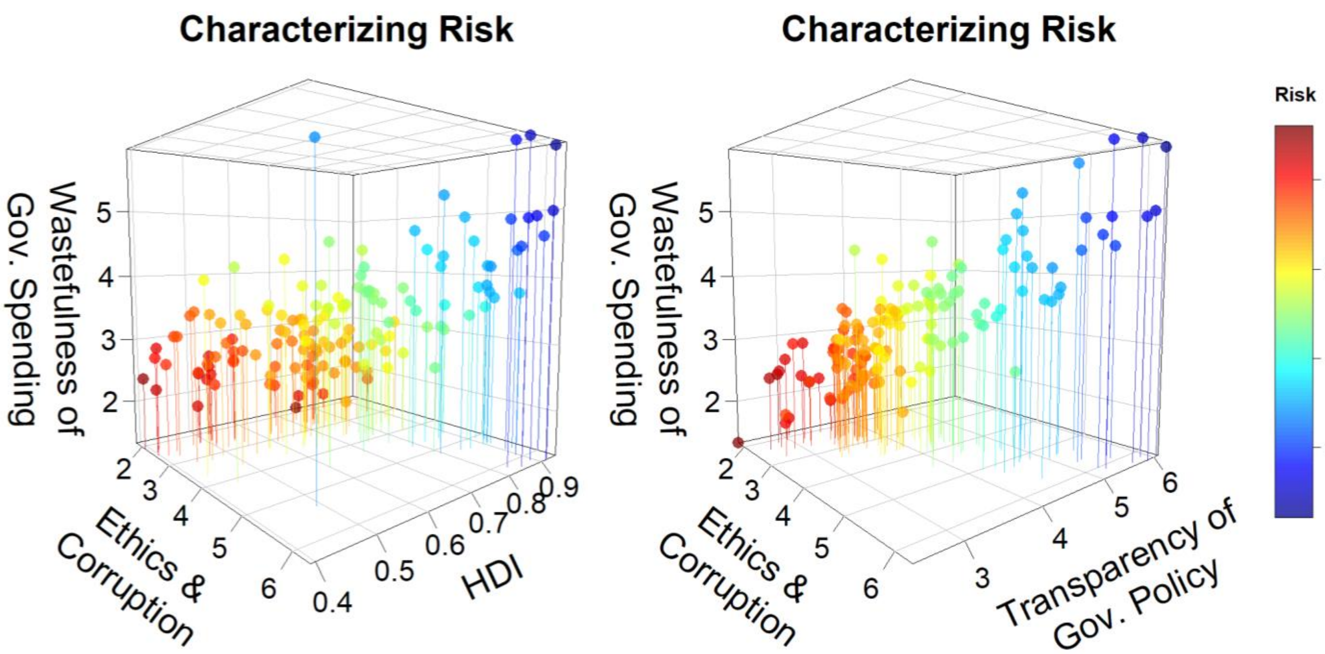
The final set of features used to create a single composite index is given below, along with their principal component loadings. The individual variances that the first principal component captures is 0.5250587^2 = 27.6% of Transparency of government policy making, 31.9% of Ethics and corruption, 14.9% of HDI, and 25.7% of Wastefulness of government spending. Overall, the first principal component captures 73.1% of the explained variance for the final set of features below.

Feature	PC1	PC2	PC3	PC4
Transparency of government policymaking	0.5250587	0.1994906	0.789052	-0.2488247
Ethics and corruption	0.5643287	0.01880304	-0.1229105	0.8161327
HDI	0.3853966	-0.87620998	-0.1182127	-0.2641048
Wastefulness of government spending	0.5072631	0.438299	-0.5901828	-0.4497359

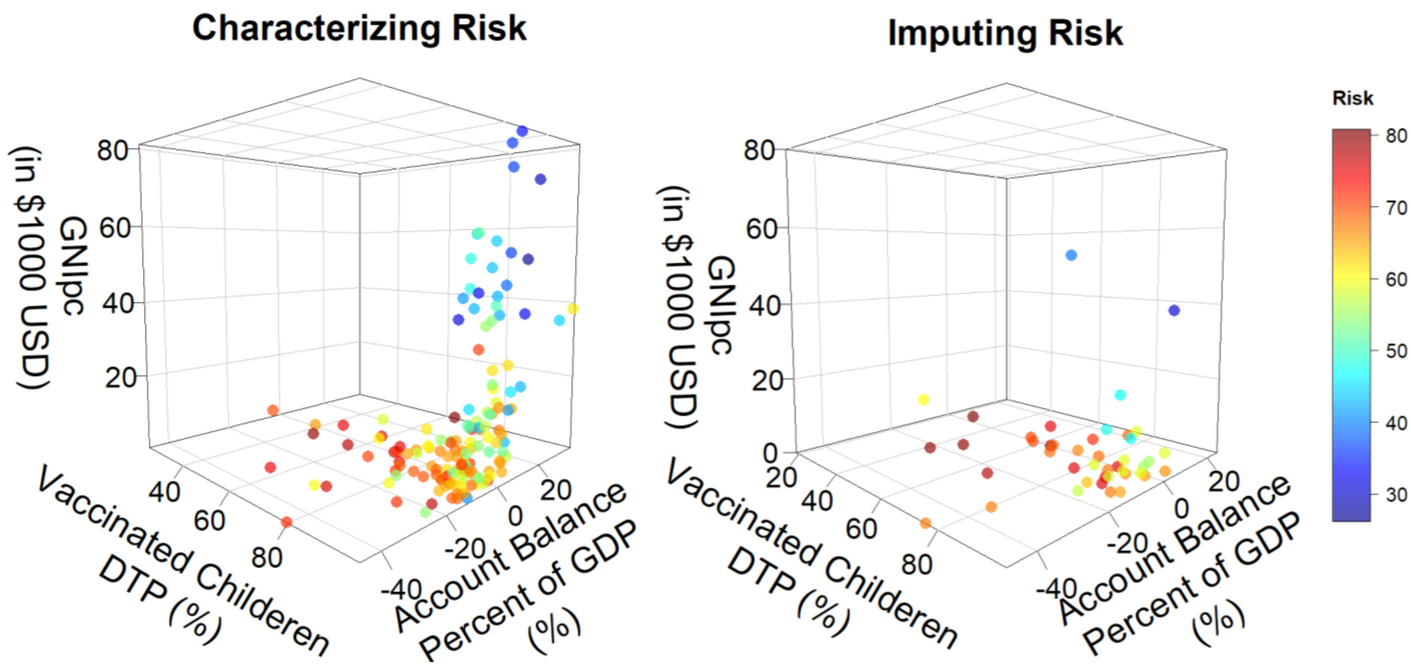
The correlation matrix below shows how the first principal component changes in value with respect to the final set of features. These correlations are used to determine how to rescale the first principal component to be interpreted as risk measurement. Ethics and corruption is a measurement that takes on values between 1 and 7 where 1 indicates poor performance and 7 indicates great performance. Transparency of government policymaking and Wastefulness of government spending follow the same scale of measure as Ethics and corruption. HDI takes on values between 0 and 1 where 0 indicates poor performance and 1 indicates great performance. Because the first principal component is positively correlated with each of the final set of features, this means that low values indicate poor performance and high values indicate great performance. Therefore, the risk index will be the negative of the first principal component and then rescaled to take on values between 0 and 100 such that higher values indicate more risk and lower values indicate less risk. This is what leads to the risk index being negatively correlated with all of the other features in the correlation matrix below. A notable observation is that the risk index maintains a strong correlated relationship with both HDI and Wastefulness of government spending, despite HDI and Wastefulness of government spending not having as strong of a correlated relationship with each other.



The risk index is exponentially averaged across the years of 2006 to 2016 for 147 countries such that the final risk for each country is weighted most heavily by the most recent years. The cubic plots below illustrate how the risk index changes in values across the final set of features, where each data point represents a single country. Both plots have drop down lines for each data point to create a shadow on the bottom pane for highlighting the shape of the feature space. Both plots show a smooth change in risk value as you move along the trend of the feature spaces. The plot on the left shows a quadractic relationship in the feature space, where the apex of the parabolic bend is positioned in the back corner. This nonlinear behavior is likely due to the presence of HDI because it had the smallest loading for the first principal component, and the second principal component captures 76.8% of the variance in HDI. The plot on the right shows a steady linear relationship across all three features. This steady linear relationship makes sense as these three features all had similar and larger loadings for the first principal component.



Recall that the data used to create the risk index was applicable to 147 out of the 194 countries of interest. Some macro-economic features are reported on consistently by most countries each year. Three features that 191 of the 194 countries have reported on are GNIpc, Vaccination Rate for DTP, and Account Balance as a Percentage of GDP. These three features are shown below in the cubic plots. The plot on the left shows the relationship between the risk index and the three well reported features for the 147 countries that have a risk measurement. This feature space shows a somewhat smooth change in values of risk, higher risk values tend to stay close to the floor where as lower risk values stay close to the open wall on the right. The plot on the right shows imputed values of risk for the remaining 44 countries by using chained random forests. Chained random forests were chosen for imputation because the plot on the left looks as though most data points of similar risk values could be grouped together by placing cuts throughout the feature space. The plot on the right shows similar behavior as that on the left, and also shows that lower income countries are the majority of the 44 countries that required imputed risk values.



The table below shows the relationship that the risk index has with respect to features of interest. The columns Q5 to Q1 represent 5 quantiles such that 20% of the observations (~38 countries) are in each quantile. Q1 represents risk values between the 1st and 20th percentiles of risk, Q2 represents the 21st and 40th percentiles, etc. The feature reporting rates show what percent of countries in each quantile reported on the features. The values for each feature (GNIpc, HDI, etc.) in this table are averages of the countries in each quantile. Across all features there is a general trend from Q5 to Q1, where risk is expected to decrease and the other features are expected to improve.

Feature	Q5	Q4	Q3	Q2	Q1
Risk	80.21	70.27	64.07	54.17	31.48
Number of Countries	38	38	38	38	39
Feature Reporting Rate	73.68%	76.32%	68.42%	78.95%	87.18%
GNIpc	$3,042.80	$4,556.70	$7,499.36	$9,106.93	$41,195.04
Account balance Percent of GDP	-8.32%	-3.82%	-5.14%	-8.21%	3.02%
HDI	0.57	0.66	0.71	0.73	0.86
Birth Cohort Mortality	4.28%	2.74%	2.19%	1.84%	0.68%
Life Expectancy	65.35	69.51	71.38	73.40	78.81
Immunization DPT	74.25%	86.60%	89.82%	93.99%	95.38%
Health expenditure public Percent of total health expenditure	44.06%	56.69%	59.55%	58.10%	70.98%
Transparency of government policymaking	3.30	3.72	3.86	4.26	5.01
Trustworthiness and confidence	3.70	4.28	4.44	4.49	5.26
Ethics and corruption	2.40	2.82	3.19	3.72	5.24
Wastefulness of government spending	2.37	2.69	2.98	3.56	4.32

REFERENCES

[1] Abeyasekera, Savitri. “Multivariate Methods for Index Construction.” Household Surveys in Developing and Transition Countries: Design, Implementation and Analysis, unstats.un.org/unsd/hhsurveys/finalpublication/ch18fin3.pdf.

[2] Lalloué, Benoît. “A Statistical Procedure to Create a Neighborhood Socioeconomic Index for Health Inequalities Analysis.” *International Journal for Equity in Health*, www.ncbi.nlm.nih.gov/pmc/articles/PMC3621558/pdf/1475-9276-12-21.pdf.

[3] Mazziotta, Matteo. “METHODS FOR CONSTRUCTING COMPOSITE INDICES: ONE FOR ALL OR ALL FOR ONE?” Rivista Italiana Di Economia Demografia e Statistica, www.istat.it/en/files/2013/12/Rivista2013_Mazziotta_Pareto.pdf.

[4] Atwater, Craig. Measuring Supply Chain Risk: Predicting Motor Carriers’ Ability to Withstand Disruptive Environmental Change Using Conjoint Analysis. Elsevier, 13 Oct. 2014, ac.els-cdn.com/S0968090X14002563/1-s2.0-S0968090X14002563-main.pdf?_tid=a07b5ef7-5d17-4bd7-b957-90e428f6ea44&acdnat=1536639108_4f3303cbd2b748887a8bdaad37a91a01.

[5] Ignatius, Joshua. A Fuzzy Decision Support System for Credit Scoring. The Natural Computing Applications Forum, 26 Sept. 2016, link.springer.com/content/pdf/10.1007%2Fs00521-016-2592-1.pdf.

[6] Malekipirbazari, Milad. Risk Assessment in Social Lending via Random Forests. Elsevier, 14 Feb. 2015, ac.els-cdn.com/S0957417415000937/1-s2.0-S0957417415000937-main.pdf?_tid=b5b7bb10-fc1f-4304-a901-b48fb39b5ec6&acdnat=1536639137_d15ea81dfedae79d6e9a9a134b19f513.

[7] Bhardwaj, Geetesh. “Credit Scoring and Loan Default.” Freshwater Biology, International Review of Finance, 30 Apr. 2015, onlinelibrary.wiley.com/doi/epdf/10.1111/irfi.12048.

[8] Arya, Shweta. Anatomy of the Credit Score. Elsevier, 12 June 2011, ac.els-cdn.com/S0167268111001259/1-s2.0-S0167268111001259-main.pdf?_tid=3902bd4e-3174-4bdb-9a84-de91ed246732&acdnat=1536639164_e341689ac4a2146efdb72517e88be942.