How does corruption affect economic growth in developing countries? A machine learning approach

Methods and Results

Qian Zhang*

May 27, 2020

Abstract

This paper intends to uncover the impact of corruption on the economic growth across the developing countries. To deal with the potential inconsistency of current corruption indicators, this paper firstly uses Support Vector Machines (SVM) to reconstruct the indicator in a continuous series on the interval of [0, 1] for 140 countries all over the world from 1996 to 2016. With this newly-constructed corruption indicator, and by Dynamic Panel Data (DPD) model and two-step "difference" GMM estimations, this paper finds that there exists a significant downward quadratic relation between corruption and economic growth, and an optimal corruption level among developing countries. Also, poorer regulator quality would enlarge the negative effect of corruption on economic growth.

keywords: Corruption, Economic Growth, SVM.

*MACSS program, University of Chicago, timqzhang@uchicago.edu.

1 Reconstruction of corruption indicator

1.1 Theoretical framework of SVM

Inspired by Gründler and Krieger (2016), this paper will mainly consider the method of Support Vector Machines (SVM), in particular the Support Vector Regression to predict continuous values. While this paper mainly focuses on the group of developing countries, the reconstruction of corruption index will cover all countries available to make accuracy approximations. The degree of corruption $c_{i,t} \in \mathcal{C} \subseteq \mathbb{R}$ for certain country i at period t can be expressed as a function $\mathscr{F} : \mathcal{X} \subseteq \mathbb{R}^m \to \mathcal{C} \subseteq \mathbb{R}$ of the extent to which the country-year pairs satisfy certain given conditions $\mathcal{X}$, where m denotes the number of conditions selected. Therefore, the degree of corruption could be expressed as

$$c_{i,t} = \mathscr{F}(x^1_{i,t}, \dots, x^m_{i,t}), \forall (i,t).$$

This paper also wants to mention that because of some unobserved features and measurement errors, it is not feasible to provide a perfect fit, and the purpose of using Support Vector Regression is to compute a function that could greatly approximate the "true" function without losing essential information.

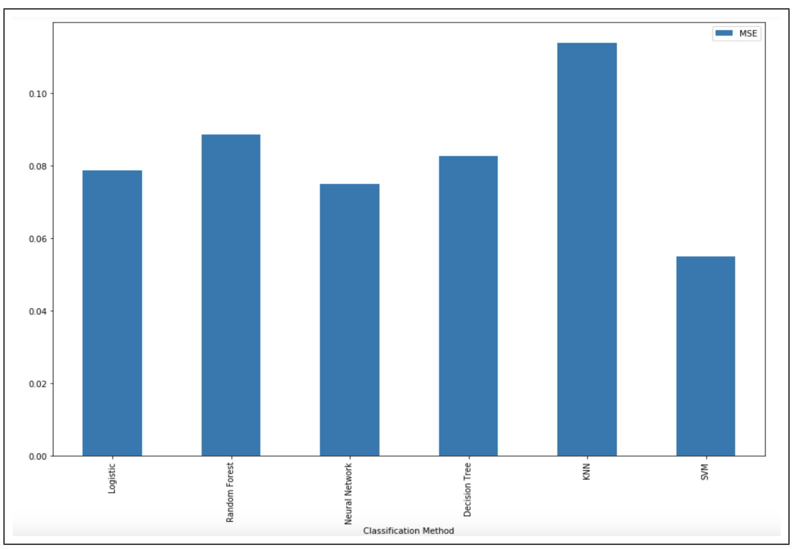
1.2 Algorithms of reconstruction

The logic of algorithm on reconstruction also mainly refers to Gründler and Krieger (2016). Firstly, some variables indicating features and characteristics need to be selected. From Graycar and Smith (2013), countries' political institutions and extent of law implementation should be considered in the measurement. Andersson and Heywood (2009) point out that culture patterns like the religious power in countries has strong relation with the perception of corruption. Also, socioeconomic conditions, internal conflicts and bureaucratic efficiency need to be included as factors suggested by Heywood and Rose (2013) to represent overall stability and governance in countries. Overall, this paper will take these six variables to form the feature space.

Secondly, this paper constructs a subset of country-year pairs from the whole data set that consists of the elements that could be identified unambiguously as being corruptive and non-corruptive. As there are several existing corruption indices, this paper mainly uses the corruption index from International Country Risk Guide (ICRG) in the model fitting and training as suggested by Treisman (2007), since the measurement methodologies of other indices have once been changed and are not suitable for cross-country panel data analysis. This corruption index is coded continuously from 0 (most corruptive) to 6 (least corruptive), and this paper inverses this index, so that higher the index, more corruptive the countries are. Then they are divided evenly in three parts. For country-year pairs with index larger than 4, this paper marks them as being corruptive (code with 1), and those with index less than 2 are coded with 0 as being not corruptive. This paper keeps those country-year pairs with coding of 0 and 1 to form a subset for later on modeling.

For the third step, this paper randomly splits this subset to create the training set and uses it to fit the SVM model, approximating the function $\mathscr{F}$ mentioned above. To further justify the SVM model, this paper controls all other conditions i.e. random seed, and run some other machine learning methods. The comparison of MSE on each method is shown in Figure 1, and it shows that the usage of SVM is statistically convincing.

Figure 1: Comparison of MSE

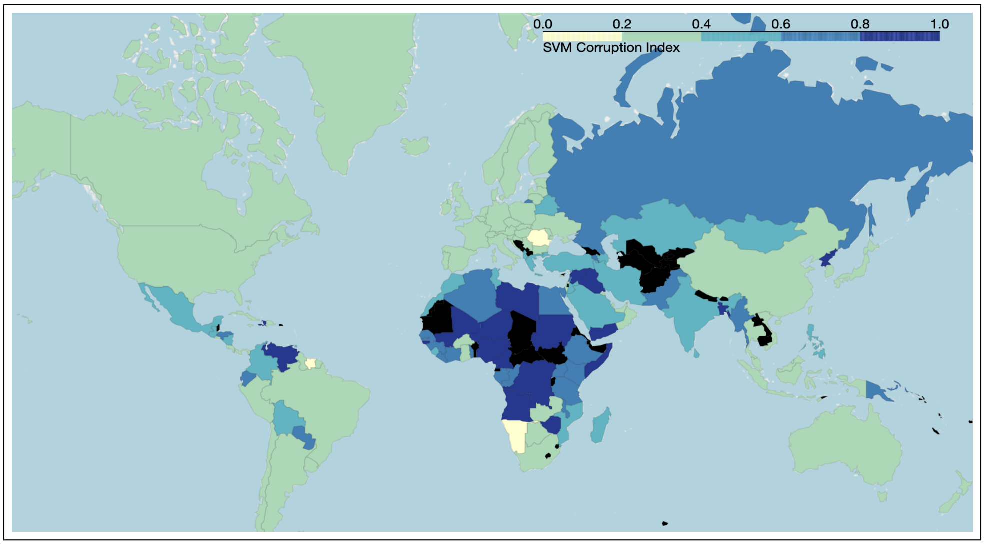


Lastly, this paper uses the estimated $\mathscr{F}$ to apply to all country-year pairs, and calculates the new corruption index $c_{i,t} \in [0, 1]$. To have a robust calculation, the bootstrapping is used.

1.3 Overview of corruption level around the world

The reconstruction of corruption index covers 140 countries from 1996 to 2016 due to the limitation of data available. Figure 2 visualizes an overview of the corruption level around the world in 2016. Deeper the color, more corruptive the country is, and countries marked black are the ones that the index does not cover. It shows that most countries in Europe, North and South America, and

Figure 2: Corruption around the world, 2016



Australia have relatively low level of corruption. It is also interesting to see that China in 2016 has a relatively low corruption level, probably because of the anti-corruption movement starting from 2012. Russia, Middle East regions and most African countries suffer from relatively high level of corruption. It could be also seen that one country's corruption level is probably related to its neighbor countries.

2 Model specification and Data

2.1 Model specification

A Dynamic Panel Data (DPD) model is used to explore the effect of corruption on economic growth among developing countries as suggested by Sharma and Mitra (2019). The DPD model mainly includes the lagged dependent variable as the explanatory variable to specify the unobserved panel effects. The baseline model is shown as

$$y_{i,t} = \beta y_{i,t-1} + \gamma c_{i,t} + \theta \mathbf{X}_{i,t} + \mu_i + \epsilon_{i,t},$$

where $y_{i,t}$ denotes the log form of GDP per capita, $c_{i,t}$ denotes the corruption indicator, $\mathbf{X}$ denotes all other control variables, and μ_i denotes country's fixed effect.

As for estimation method, Arellano and Bond (1991) point out the "difference" GMM estimation, in which the lagged dependent variable serves as the instrument and all variables are taken the first difference. After taking the first difference, the fixed effect μ_i will be wiped out. Blundell and Bond (1998) also propose a "system" GMM estimation, which aims to alleviate the problem of poor instruments by adding extra moment conditions. However, as mentioned by Roodman (2009), while both "difference" and "system" GMM have the estimation problem from instrument proliferation, such problem is particularly dangerous in "system" GMM, which could cause severe biases. Therefore, considering the data set and number of instruments that would be used, this paper will mainly consider the "difference" GMM for estimation.

2.2 Data and Variables

The data mainly comes from World Development Indicators (WDI) of World Bank, from 1996 to 2016 over 128 developing countries. The dependent variable is the GDP per capita and will be in the log form. The variable of corruption is directly from the reconstructed corruption index. The index of Regulatory Quality is also included suggested by Dzhumashev (2014), which comes from ICRG with continuous values, and is transformed to the inversed value, i.e. higher the index, poorer the regulator quality of one country. This paper also includes some important control variables to capture the aspects of trade openness, government consumption, capital investment, natural resources, inflow of foreign direct investment (FDI), and inflation from WDI. The descriptive statistics is shown in Table 1.

Table 1: Summary of descriptive statistics

Variables	Obs	Mean	Std.Dev.	Min	Max
GDP PER CAPITA	2630	5448.30	8370.52	187.52	64864.71
CORRUPTION	1974	0.50	0.22	0.00	1.00
REGULATORY QUALITY	1948	2.21	0.86	0.00	4.00
INFLATION	2623	13.44	111.67	-36.57	4800.53
GOVERNMENT CONSUMPTION	2464	14.66	7.21	0.91	135.81
TRADE OPENNESS	2553	82.55	50.86	0.03	442.62
CAPITAL INVESTMENT	2464	22.51	8.23	-2.42	69.67
FDI INFLOW	2605	4.38	7.46	-37.16	161.82
NATURAL RESOURCE	2630	10.19	12.85	0.00	86.45

Note: GDP per capita is in constant 2010 US dollar. Inflation is in the percentage form. Government Consumption, Trade Openness, Capital Investment, FDI Inflow and Natural Resource are all in the form of percentage share of GDP. Particularly, Trade Openness is calculated by (imports + exports) / GDP, where imports and exports are the imports and exports of goods and services.

3 Empirical result and Discussion

3.1 Empirical result

Table 2: Effect of corruption on growth, two-step "difference" GMM

Dependent variable: LGDP	(1)	(2)	(3)	(4)
LAGGED LGDP	0.974***	0.978***	0.973***	0.972***
	(0.002)	(0.002)	(0.003)	(0.004)
CORRUPTION	-0.086***	0.010	-0.009	0.061**
	(0.007)	(0.020)	(0.023)	(0.030)
CORRUPTION^2		-0.092***	-0.069***	-0.039*
		(0.017)	(0.020)	(0.023)
REG_QUALITY			-0.006*	0.013***
			(0.003)	(0.005)
REG_QUAL×CORR				-0.039***
				(0.009)
Observations	1,738	1,738	1,728	1,728
Number of countries	94	94	94	94
Serial Correlation Test (p value)	0.615	0.615	0.519	0.566
Sargan Test (p value)	0.656	0.555	0.582	0.583

[1] Control Variables: Inflation, Government Consumption, Trade Openness, Capital Investment, FDI Inflow, Natural Resources.

[2] LGDP is denoted as the log form of GDP per capita.

[3] p values in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

This paper mainly uses two-step "difference" GMM estimation to get the robust estimates. Three hypotheses are going to be tested. (1) High level of corruption has a negative impact on the economic growth among developing countries. (2) Poorer regulatory quality has a negative impact on economic growth among developing countries. (3) Poorer regulatory quality will enlarge the negative effect of corruption on economic growth among developing countries. The empirical result is shown in Table 2. The first column shows the baseline model. The second column adds the squared corruption index to further check if there exists a quadratic relation between corruption and economic growth, and the first two columns will test Hypothesis 1. The third column adds the regulatory quality to test Hypothesis 2. The fourth column further adds the interaction of regulatory quality and corruption to test Hypothesis 3.

3.2 Discussion

Hypothesis 1: High level of corruption has a negative impact on growth.

This hypothesis echoes the "sand the wheel" hypothesis pointed out by Mauro (1995). From the baseline model, the coefficient of corruption is significantly negative at 1% level. It verifies the hypothesis and suggests that with all other variables fixed, higher corruption level will negatively affect economic growth in developing countries. Moreover, an interesting result is found when adding the squared corruption term, and the coefficient of corruption goes positive and the coefficient of squared term is significantly negative. Such a relation keeps significant at column 4. It suggests that with all other variables equal, there exists a downward quadratic relation between corruption and economic growth, and an optimal level of corruption exists among developing countries. In other words, when corruption level grows higher, it will firstly positively affect growth, and after the optimal level, it will have negative impact. Such result echoes the hypothesis and empirical evidence from Acemoglu and Verdier (1998). Also, the significantly positive coefficient of lagged LGDP shows that the past levels of corruption have an impact on both current and future economic growth.

Hypothesis 2: Poorer regulatory quality has a negative impact on growth.

Kinda et al. (2009) suggest that the poorer regulatory quality of government will impede the growth in developing countries. From column 3, the coefficient of regulatory quality is negative and in general justifies the hypothesis. However, the coefficient is just significant at 10% level, and some empirical evidence (Gani, 2011) shows a negative but insignificant impact of regulatory quality on growth. Therefore, further analysis on the impact of regulatory quality to economic growth could be done, but it is irrelevant to the major topic of this paper.

Hypothesis 3: Poorer regulatory quality will enlarge the negative effect of corruption on growth.

Aidt et al. (2008) develop a theory and provide the empirical evidence that the relationship between corruption and growth is related to the quality of regulation. All coefficients of column 4 are significant at least on 10% level, and the overall marginal effect of corruption is $(0.061 - 0.078 \times \text{CORR} - 0.039\times \text{REG\_QUAL})$. Given a high level corruption $(0.061 - 0.078 \times \text{CORR} < 0)$ and all other variables fixed, if regulatory quality goes poorer, i.e. one unit increase of regulatory quality index, the overall negative effect of corruption will become larger, which verifies the hypothesis.

References

Acemoglu, Daron and Thierry Verdier, "Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach," *The Economic Journal*, dec 1998, *108* (450), 1381–1403.

Aidt, Toke, Jayasri Dutta, and Vania Sena, "Governance regimes, corruption and growth: Theory and evidence," *Journal of Comparative Economics*, 2008, *36* (2), 195–220.

Andersson, Staffan and Paul M. Heywood, "The politics of perception: Use and abuse of transparency international's approach to measuring corruption," *Political Studies*, 2009, *57* (4), 746–767.

Arellano, Manuel and Stephen Bond, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, 1991, *58* (2), 277–297.

Blundell, Richard and Stephen Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 1998, *87* (1), 115–143.

Dzhumashev, Ratbek, "Corruption and growth: The role of governance, public spending, and economic development," *Economic Modelling*, 2014, *37*, 202–215.

Gani, Azmat, "Governance and Growth in Developing Countries," *Journal of Economic Issues*, mar 2011, *45* (1), 19–40.

Graycar, Adam and Russell G. Smith, "03_2011_Heinrich_Hodess_Measuring Corruption.pdf," in "Handbook of global research and practice in corruption" 2013, pp. 18–33.

Gründler, Klaus and Tommy Krieger, "Democracy and growth: Evidence from a machine learning indicator," *European Journal of Political Economy*, 2016, *45*, 85–107.

Heywood, Paul M. and Jonathan Rose, "Close but no Cigar: The measurement of corruption," *Journal of Public Policy*, 2013, *34* (3), 507–529.

Kinda, Tidiane, Patrick Plane, and Marie-Ange Véganzonès-Varoudakis, "Firms' productive performance and the investment climate in developing economies: an application to MENA manufacturing," *World Bank Policy Research Working Paper Series*, 2009, *4869* (March).

Mauro, Paolo, "Corruption and Growth," *The Quarterly Journal of Economics*, aug 1995, *110* (3), 681–712.

Roodman, David, "Practitioners' corner: A note on the theme of too many instruments," *Oxford Bulletin of Economics and Statistics*, 2009, *71* (1), 135–158.

Sharma, Chandan and Arup Mitra, "Corruption and Economic Growth: Some New Empirical Evidence from a Global Sample," *Journal of International Development*, 2019, *31*, 691–719.

Treisman, Daniel, "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?," *Annual Review of Political Science*, 2007, *10*, 211–244.