NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Anastasiia Shukhova,
Yulii Nisnevich

MEASUREMENT OF VALIDITY OF CORRUPTION INDICES

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: POLITICAL SCIENCE
WP BRP 42/PS/2017

Anastasiia Shukhova[1], *Yulii Nisnevich*[2]

MEASUREMENT OF VALIDITY OF CORRUPTION INDICES[3]

Corruption is an intrinsically latent phenomenon, which makes it a challenging task to measure it and requires the use of indirect indicators. The academic community and non-government organizations have proposed various indices that differ in terms of their methodology, data and coverage. In this paper, we estimate construct validity of the most widely used indices of corruption: The Corruption Perceptions Index, The Control of Corruption Index, The Bribing and Corruption Index, The Corruption Index, and The Rule of Law: Absence of Corruption. In this paper we show that Corruption Index by the International Country Risk Guide and Absence of Corruption Index are not constructively valid and, therefore, are not suitable for the use in scholarly research. We also show that all indices provide poor estimates of a corruption level in the highly corrupted group of countries.

JEL classification: C01, C23, P50

Keywords: corruption index, validity

---

[1] University of Mannheim, MA Political Science student; E-mail: aashukhov@mail.uni-mannheim.de

[2] National Research University Higher School of Economics. Professor, Leading Research Fellow; E-mail: jnisnevich@hse.ru

[3] The authors thank Denis Stukal for his very useful comments and helpful assistance in the research.

Introduction

Corruption is arguably a major problem in developing and some of the developed countries. It not only imposes a tax on public services and private sector activities, but also creates potentially severe efficiency loses for the economy in general (Krueger, 1974; Shleifer & Vishny, 1993). In addition, corruption is a multifold phenomenon that encompasses different kinds of relationships in a number of spheres, including legislation, policy implementation, law enforcement, etc.

Corruption is an intrinsically latent phenomenon, therefore in order to measure it one has to use indirect data sources. Current measurements of corruption are based on two types of data -- expert and public opinion surveys -- that both have certain flaws and biases. First, people can misreport their true beliefs about their personal experience or perception of corruption. This can be especially common for non-democratic regimes, where respondents may falsify their preferences (Kuran, 1995) and dissemble the truth because of being afraid of the possible punishment from the government (Philp, 2006). But even in democratic regimes some rate of misreport can be due to social desirability bias preventing respondent to freely acknowledge the fact of having been involved in corruption. Second, in different cultural contexts, people can have distinct perceptions of what corruption is. As a result, the embezzlement of public funds can provoke a tremendous public scandal in one country, but be considered a daily and even socially acceptable routine in another (Smith, 2015). Finally, individual perceptions of the level of corruption can be fallacious, because of the lack of specialized or complete knowledge of the ways politics in done.

Due to the pitfalls of different types of data and the absence of an ideal indicator of corruption, both the academic community and non-government organizations have proposed various indices that differ in terms of their methodology, data and coverage. Unsurprisingly, these indices can produce dissimilar estimates of the level of corruption in a given country depending on the sources of the data and the methods used for their aggregation. As an example, we show in Figure 1 dynamics of the level of corruption in two countries, Latvia and Madagascar, produced by two different indices of corruption, Corruption Perceptions Index by Transparency International and World Governance Indicators by the World Bank. Both indices are scaled to range between 0 and 1, where 1 represents the maximum level of corruption.

In the case of Latvia, one can see from Figure 1 both a large difference in the reported levels of corruption in 2006 and 2011, and different dynamics between 2006 and 2008. Importantly, in January 2006 Latvia ratified the UN Convention against Corruption, and soon enough, in March 2006, passed the Law on Public Procurement that was supposed to increase the transparency of public procurement procedures (Corruption Prevention and Combating Bureau, 2006). However, the two indices lead to different, and even opposite, conclusions about the effects of the law, an undesirable result for policy evaluation purposes.

One can see even more striking dissimilarity in the case of Madagascar. According to CPI, corruption was abundant and flourishing in Madagascar in 2002 (0.83), although it was quite mild according to WGI (0.48). These drastic differences may be due to these indices' distinct reaction to the violent political crisis in Madagascar after the 2001 presidential elections. However, the dynamics of corruption is also dissimilar, which opens the question about which index provides a better and more accurate representation of the situation with corruption in the country, which is a question about measurement validity

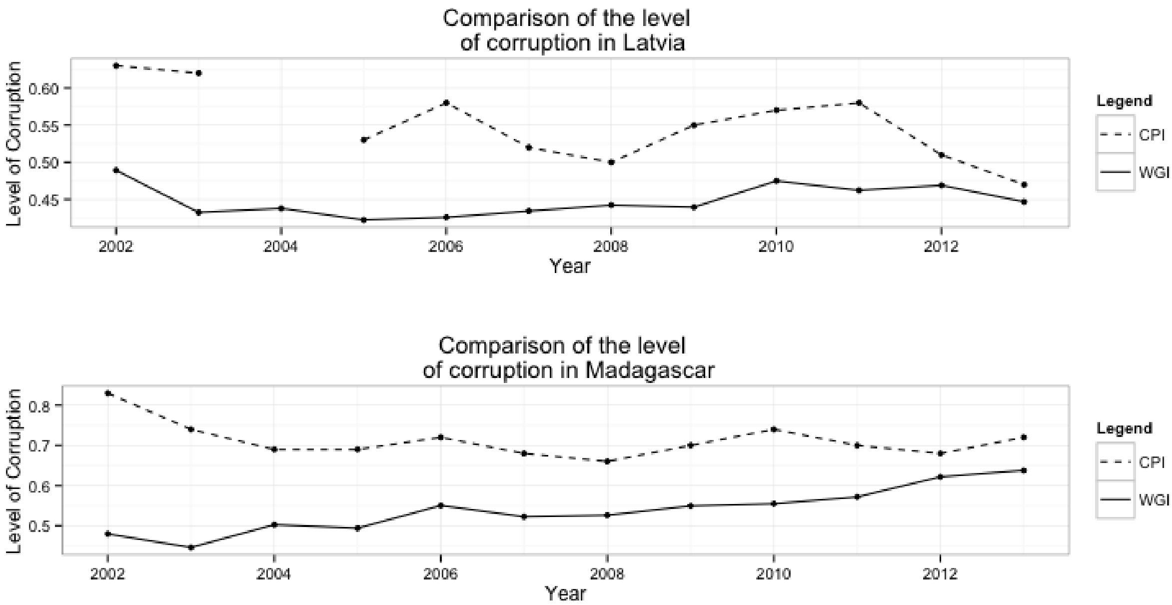

Figure 1. The level of corruption in Latvia and Madagascar over 10 years estimated by Corruption Perception Index and Control of Corruption Index in World Government Indicators.
Note: In Latvia, CPI has not provided estimates for the level of corruption from 2003 to 2005.

Although validity has for a long time been a big concern among psychologists and has recently attained attention in the political science community, there is scarce systematic research on the validity of corruption indices.

Wilhelm (2002) is one of the first seminal papers solely aimed at analyzing construct validity of corruption indices. Even though Wilhelm (2002) presents clear methodology for analyzing construct validity using indices of activity on the black market and excessive regulation, the use of these variables limits construct validity research, since these variables represent only small part of corruption activity that indices claim to measure. Ko & Samajdar (2010) present the next systematic and comprehensive analysis of several corruption indices that is of big interest for scholarly research. However, the paper uses only correlation analysis, which is not enough for the validity estimation (Cronbach & Meehl, 1955). Nisnevich & Stukal (2012) deal with this problem using wide range of statistical methods, although they use only three indices of corruption. In comparison to previous research in this field, our paper presents comprehensive research that involves wide range of corruption indices under analysis and statistical methods to test their construct validity.

In this paper, we estimate construct validity of the most widely used indices of corruption: The Corruption Perceptions Index – CPI – (Transparency International), The Control of Corruption Index – WGI – (World Government Indicators), The Bribing and Corruption Index – BI – (The International Institute for Management Development), The Corruption Index – ICRG – (International Country Risk Guide), and The Rule of Law: Absence of Corruption – WJP – (World Justice Project).

In order to measure the validity of the corruption indices, we implement partial correlation analysis, principal component analysis, factor and regression analysis. In this paper we show that Corruption Index of the International Country Risk Guide and Absence of Corruption Index are not constructively valid and, therefore, are not suitable for the use in scholarly research. We also show that all indices provide poor estimates of a corruption level in the highly corrupted group of countries. On the basis of our analysis, we can say that in a research one may use CPI and WGI in the lowly corrupted groups, and with some caution in the highly corrupted groups.

The rest of the paper is organized as follows. First, we discuss the notion of corruption and provide some history of the measurements of corruption. Then we discuss the methodology of indices in detail to identify strong and weak sides of the indices before we start analyzing it quantitatively. Next, we discuss our validation methodology and justify our methodology; afterwards we examine the specifics of our data. Finally, we conduct analysis and discuss our results.

Corruption: Theory and Measurement

History of corruption estimation

In 1984, Political Risk Services Group was the first organization to present systematic cross-country measurements of the level of corruption in International Country Risk Guide (ICRG) across 146 countries. Next, since 1989, International Institute for Management Development (IMD) has been doing the world competitiveness report, in which IMD estimated competitiveness using multiple measurements of the business activity and quality of government, the latter included estimates of the level of corruption. In 1995 and 1996, Transparency International and World Bank respectively began to estimate the level of corruption almost all over the world. Finally, the World Justice Project was the latest project to estimate the level of corruption across multiple countries since 2010.

A short overview of the methodologies of the indices and their pitfalls.

Before we start to implement a quantitative analysis of the construct validity of indices, it seems reasonable first to look at the indices themselves and to figure out what aspects of corruption they are intended to measure and how indices are constructed. Unfortunately, it is not feasible to do for all of the indices we are working with, since for the corruption index (ICRG) methodological description is not available for free use.

The Corruption Index (International Country Risk Guide)

Since 1984 one of the first sources of corruption level estimation, International Country Risk Guide, has been presenting data on the financial, economic, and political risks across different countries. Within an index of the political risk, ICRG presents a corruption component that ranges from 0 to 6, where 0 corresponds to the highest possible level of corruption, 6 – to the lowest possible level of corruption.

Officially, the Corruption Index captures the following spheres of illegal activity: ‘‘actual or potential corruption in the form of excessive patronage, nepotism, job reservations, ‘favor-for-favors’, secret party funding, and suspiciously close ties between politics and business’’ (Political Risk Survey Group, 2014: 4-5).

Galtung (2005) notes that the index developers base the estimation of the level of corruption using the data on the longevity of the rule and the way a government came to power. Here, the theoretical mechanism is the following: the longer the government stays in power, the higher is the chance of the presence and abundance of the patronage, nepotism and other corrupt activities that ICRG uses for the index construction. Williams & Siddique (2008) state that the measure of the corruption based on such considerations is too indirect and imprecise to use it as the estimate of the level of corruption in the research. For example, in some corruption abundant countries, e.g. Egypt or Russia, government stays in power for a restricted amount of time and changes regularly, however, this does not imply the absence of nepotism and corrupt activities within the government. Williams & Siddique (2008) also point out that ICRG represents only political risk associated with corruption, which makes it unreasonable to use this index as a measure of corruption. However, neither Williams & Siddique (2008), nor Galtung (2005) have empirically checked their claims.

The Bribing and Corruption Index – BI – (The International Institute for Management Development)

The International Institute for Management Development (IMD) has been conducting a competitiveness research since 1989. The World Competitiveness Yearbook presents 327 economic and political measurements, including the level of corruption. In The World Competitiveness Yearbook, IMD measures the amount of bribery and corruption across countries. IMD constructs The Bribery Index using one question in the expert survey, where respondents should assess the existence of the bribery and corruption in a country on 0-10 scale (IMD Competitiveness Yearbook, 2014). Then IMD averages respondents' assessments by country and presents these means as final values of The Bribery Index.

The Corruption Perceptions Index – CPI – (Transparency International)

Since 1995 Transparency International has been publishing Corruption Perceptions Index (CPI). CPI is an aggregate index that ‘‘ranks countries in terms of the degree to which corruption is perceived to exist among public officials and politicians.’’ (Transparency International, 2011: 1) Surveys and assessments of the level of corruption should include ``questions relating to the bribery of public officials, kickbacks in public procurement, embezzlement of public funds, and questions that probe the strength and effectiveness of public-sector anti-corruption efforts.’’(Transparency International, 2011: 4) CPI data sources include only expert surveys and assessments.

In 2012, Transparency International has drastically changed the methodology of the index that makes it impossible to compare the values of the index before and after the methodological change. We therefore use in our research CPI data before 2012, whereas the values after 2012 we imputed with the country specific average of the index.

Transparency International builds the index from several sources, whereas each of them should meet the following requirements. First, a source should measure the level of corruption in the public sphere and among public officials. Second, a source should provide cross-country estimates and should have methodology that is the same for every country. A country receives its corruption score only if there are at least three data sources for this country. Usually, very small and poor countries do not have three sources of information, which causes the existence of the Missing At Random data in CPI.

Transparency International calculates CPI using the methods of ``matching percentiles” and beta-transformation to standardize the estimates of corruption that come from different sources. The mean of standardized measurements of corruption for every country is the final value of CPI.

The Control of Corruption Index – WGI – (World Government Indicators)

In 1996, World Bank introduced first data on the World Government Indicators (WGI). WGI consists of 6 indicators measuring the quality of government, one of which is Control of Corruption Index. This corruption index captures "perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests." (Kauffman et al., 2011: 4)

The Control of Corruption Index is a composite measurement of corruption. Even though Control of Corruption Index and Corruption Perception Index use some data sources that are similar for both of the indices, the main distinction of Control of Corruption Index is that it uses not only expert surveys, but also public opinion polls that may be useful for the capturing the perception of the level of corruption in everyday life.

In order to construct Control of Corruption Index, World Bank uses unobserved components model (UCM), which relies on the premise that "each of the individual data sources provides an imperfect signal of some deeper underlying notion of governance that is difficult to observe directly." (Kauffman et al., 2011: 9)

UCM models each of six governance indicators in the following way:

$$y_{jk} = a_k + b_k(g_j + e_{jk}) \qquad (1)$$

where y_{jk} stands for a data source k in country j, a_k and b_k are parameters that map unobserved governance in country *j*, g_j, to the actual data from the source y_{jk}, and e_{jk} is random error (Kaufmann et al., 2011: 9-10). Unobserved components model assumes that unobserved governance in country *j* has standard normal distribution, random error e_{jk} is distributed with standard normal distribution within a country, but has different variance across different data sources. The model also assumes that there is no covariation across the sources. Weighted mean of the rescaled sources for each country represents the final value for the Control of Corruption Index, where the smaller is the variance of the error term of the sources (in other words, the more informative the signal of the level of corruption is), the bigger the weight of this source is in the final value of the level of corruption.

While Corruption Perceptions Index has a restriction on the number of sources for every country, Control of Corruption Index measures the level of corruption for every country, for which at least one source exists. In Control of Corruption Index, standard deviation of the final measurement of corruption depends on the number of the sources, i.e. the bigger the number of the sources for country *j*, the smaller the variance of the estimated level of corruption for this country is.

However, this model has some disadvantages that come from the strictness of its assumptions. First, model assumes zero covariation of errors across countries and data sources. Second, model assumes linear association between latent governance (in this case level of corruption) and data sources.

To identify the next assumption, we need first to consider the classical linear model that we present in (2).

$$y_j = a + b * g_j + e_j \tag{2}$$

In this model, $Cov\left(y_j, g_j\right) = b*Var\left(g_j\right)$ and $b = \frac{Cov(y_j, g_j)}{Var(g_j)}$, i.e. beta represents normalized covariation between the *y* and *g*. In addition, in this model $Cov(y_j, e_j) = Var(e_j)$. In order to obtain unbiased estimation of model parameters, these assumptions should hold.

Now, go back to the model (1). Here, $Cov\left(y_j, g_j\right) = b*Var\left(g_j\right)$ and $b = \frac{Cov(y_j, g_j)}{Var(g_j)}$, but $Cov(y_j, e_j) = b*Var(e_j)$, which implies that $b = \frac{Cov(y_j, e_j)}{Var(e_j)}$. Therefore, in the UCM model a strict assumption should hold that $b = \frac{Cov(y_j, e_j)}{Var(e_j)} = \frac{Cov(y_j, g_j)}{Var(g_j)}$, which is difficult to guarantee. If this assumption does not hold, then the estimates of the level of corruption can be biased.

We have shown that unobserved components model has several strict assumptions that sometimes can be challenging to guarantee. This implies some weakness of the model underlying World Governance Indicators, including Control of Corruption Index.

The Rule of Law: Absence of Corruption – WJP – (World Justice Project)

Since 2010 World Justice Project has been working on the Rule of Law Index that consists from 9 components, including corruption. Absence of corruption component of the index considers three forms of corruption: “bribery, improper influence by public or private interests, and misappropriation of public funds or other resources (embezzlement)” (Botero & Ponce, 2011: 10). Absence of corruption component measures the level of corruption among public officials, including military and police.

World Justice Project calculates its measure of corruption using both expert and public opinion surveys. Public opinion surveys acquire information about people's perceptions and experience of corruption, openness and accountability of the government, and the extent to which society is exposed to the common crimes. World Justice Project conducts public opinion survey only in three biggest cities in a country. However, the level of corruption, as well as public perceptions about the existence of corruption, may considerably vary across cities with different sizes. Consequently, one can suggest the existence of the selection bias in the data on the level of corruption in this index.

In the expert survey respondents are the experts in one of the following spheres: civil and commercial law, criminal justice, labor law, and public health. Respondents should be either law professors, or practicing attorneys with substantial experience in at least one of four spheres. (Botero & Ponce, 2011) Comparing to the indices described above, Absence of Corruption Index does not include businessmen in their sample of respondents, even though questionnaire has some questions on how corrupted the business is. Unfortunately, legal attorneys and law professors may not have all necessary information about the way one can conduct business in a country, therefore, their answer may not represent the actual level of corruption in business activities. Consequently, a ramification of the absence of businessmen in the respondent sample may lead to biased estimates of the level of corruption in a country.

The methodology of the index is quite straightforward – it is an average of all questions in the survey, where the answers on these questions can be either 1, or 0.

Unfortunately, the country range for this index is restricted to 35 countries in 2010, and to 65 in 2012, which introduces the biggest share of the missing observations in our data.

Data and Methodology

Methodology

In this paper we analyze validity of the existent indices of corruption. In social sciences, there are four basic types of validity:

- *content validity* considers whether an index covers all aspects of the latent variable;
- *construct validity* refers to whether a high correlation among indicators of the same latent variable exists;
- *concurrent validity* refers to whether there is a high correlation with the initially valid indicator;
- *predictive validity* considers whether an indicator of interest can predict future outcomes of the latent variable.

It seems to be impossible to empirically test for concurrent and predictive validity, for these types of validity assume existence of the initially valid indicator that is absent in corruption studies. Content validity is difficult to test, since it requires a developed theory of corruption that is absent in political science. Finally, construct validity almost does not impose severe requirements that make this type of validity feasible to test; therefore, in this paper we estimate construct validity and reliability of the most popular indices of corruption.

In order to test construct validity of the corruption indices and make well-founded judgment about the validity of the indices, we apply several statistical methods to the corruption data. First, we are going to analyze partial correlation coefficients. Ideally, if all coefficients are constructively valid, then partial correlations of any two indices, while removing the effect of the third index, should be close to zero. Substantively, this means that if indices measure the same aspects of the latent variable, then removing the information of one index from the other indices leaves no systematic information in both variables, since they contain only a random error. On the contrary, if an index captures only a part of the latent variable, then there is still some information after the removal of its effect from the other variables.

After we analyze how much information does an index have relative to the other indices, we are going to implement another technique to estimate the quality of association of the indices of corruption. If all indices of corruption measure the same latent variable, then they should form one-dimensional space that itself would construct a synthetic variable representing the latent variable, corruption. If indices measure different aspects of corruption or measure not only the level of corruption, but, say, quality of government, then they are likely to form at least two-dimensional space, where every space represents a distinct and independent synthetic latent variable. The latter case implies invalidity of the indices of corruption.

In addition, we also test reliability of the corruption indices. High reliability of the index means that the variance of its stochastic component is low enough to make precise conclusion about the level of corruption. In order to understand how big are the errors of the indices, we use confirmatory factor analysis. In this case, we treat the level of corruption as a latent variable that explains some of the variation in the index of corruption. The bigger the share of the variance of the latent variable is in the variance of the index, and the smaller is its error, the more reliable is the index of corruption. Analytically, this idea may be presented in the following way:

$$CorruptionIndex_i = b * LatentCorruption + e_i \qquad (5)$$

where e_i corresponds to the error unexplained by the latent variable.

The last step of the analysis is to check the performance of the indices of corruption across countries with low and high level of corruption. If indices are valid, then correlation between them should stay the same across different types of countries. If this requirement is violated, then we can state the invalidity of indices. Different levels of association of the corruption indices across different types of countries may come from the subjectivity of the expert survey. For instance, if the country is known for its high level of corruption in the past, then an expert may unintentionally higher the level of corruption, even though at the moment the level of corruption may be lower. This mechanism also works in another direction for countries with the low level of corruption. In order to analyze the performance of the corruption indices across countries with either low or high level of corruption, we run several regressions with the interaction effect. Analytically, we are going to estimate the following model:

$$\begin{aligned} CorruptionIndex_i^{(j)} &= b_0 + b_1 CorruptionIndex_i^{(k)} + b_2 HighlyCorruptedCountry_i + \\ &+ b_3 CorruptionIndex^{(k)} * HighlyCorruptedCountry_i + e_i \end{aligned} \qquad (6)$$

where e_i corresponds to the error term, and corruption indices have different subscripts to point out that we use different indices in the regression model. The insignificance of the effect of the interaction term implies equal association of the indices across countries with different level of corruption.

Data

In the research we use the following widely used indices of corruption:

1. The Corruption Perceptions Index – CPI – (Transparency International);
2. The Control of Corruption Index – WGI – (World Government Indicators);
3. The Bribing and Corruption Index – BI – (The International Institute for Management Development);
4. The Corruption Index – ICRG – (International Country Risk Guide);
5. The Rule of Law: Absence of Corruption – WJP – (World Justice Project).

Table 1 presents the description of the data. For the sake of comparability across different indices, we rescaled all indices of corruption making them vary from 0 to 1, where 0 indicates the complete absence of corruption, and 1 refers to the absolutely high level of corruption.

Table 1. Data Description

Index	Year	Number of observations	Share of Missing Data	Mean	Standard Deviation
CPI	1997 – 2013	2390	0.29	0.51	0.20
WGI	1998 – 2013	2590	0.19	0.58	0.22
BI	1997 – 2013	694	0.79	0.54	0.27
ICRG	1997 – 2013	1285	0.60	0.46	0.20
WJP	2010 – 2013	291	0.91	0.55	0.21

Note: dataset contains data on all countries that are members of the UN. All data available at official web-sites of the organisations.

The main disadvantage of the data is the significant amount of missing data that we observe for several reasons. Firstly, initially different indices estimated level of corruption for different number of countries (if CPI and WGI estimate corruption for the most of the countries, BI provides estimates only for a limited number of countries). Secondly, in different periods indices estimated the level of corruption for different set of countries. For instance, CPI measures the level of corruption only if a country have at least three sources of corruption proxies in a given year. Finally, WJP appeared only in 2010, and this fact explains the biggest share of the missing data in this index.

Results

Partial Correlation Analysis

Partial correlation coefficient of the first order represents the correlation coefficient of two variables, while controlling for the effect of the third variable on both of the variables. Substantially this means that we take away the intermediate effect of the control variable from the correlation between the other two. In terms of corruption, we assume that if control variable provides a valid and comprehensive measurement of corruption, then controlling for its effect on the other two variables will leave no substantial information in these variables, therefore we expect partial correlation coefficients to be close to zero.

The main advantage of this method is that it does not impose strict assumptions on the number of observations, which makes it possible to use all variables in the analysis.

Table 2 presents results of the partial correlation analysis. The partial correlation coefficient of CPI and WGI controlling for the effect of BI is quite high (0.81). Since BI mostly represents the level of bribery that is only one part of corruption activity, this result speaks in favor of the construct validity of CPI and WGI, for these indices reflect the broader range of corruption activities.

Next, if we look at the partial correlation coefficient of WGI and BI controlling for the effect of CPI ($R = 0.00$) and the partial correlation coefficient of CPI and BI controlling for the effect of WGI ($R = 0.57$), we may say that CPI captures more aspects of corruption than WGI does, since deletion the intermediate effect of CPI leaves no correlation in the indices. However, this effect does not hold if we look at the significant partial correlation coefficient of WGI and WJP controlling for the effect of CPI ($R = 0.44$), and insignificant partial correlation coefficient of CPI and WJP controlling for the effect of WGI ($R = -0.12$). In this particular case, WGI captures wider range of corruption activities than CPI, which implies that WGI is more constructively valid than CPI. These quite contradictory results force us to further scrutinize the validity of CPI and WGI.

Table 2. Partial correlation coefficients

Indices	Control Variables				
	WGI	**CPI**	**BI**	**ICRG**	**WJP**
WGI, CPI			0.81*** (517)	0.90*** (819)	0.94*** (177)
CPI, BI	0.57*** (517)			0.87*** (342)	0.84*** (28)
WGI, ICRG		0.28*** (819)	0.51*** (259)		0.74*** (69)
CPI, ICRG	0.11*** (819)		0.45*** (342)		0.72*** (76)
WGI, BI		0.00 (517)		0.76 *** (259)	0.88*** (21)
WGI, WJP		0.44*** (177)	0.38 (21)	0.60*** (69)	
BI, WJP	0.05 (21)	0.33** (28)		0.77** (12)	
BI, ICRG	0.09 (259)	0.03 (342)			0.79 ** (12)
CPI, WJP	-0.12 (177)		0.18 (28)	0.52*** (76)	
ICRG, WJP	0.000 (69)	-0.14 (76)	0.25 (12)		

Note: ***** --** p < 0.01, ** -- p <0.05, * -- p<0.1. Blank cells if due to the absence of the third unique variable.

Now, consider the results of the partial correlations analysis with respect to ICRG and WJP. In general, if we look at the last two columns of Table 2, we see that all partial correlation coefficients are high and significant after controlling either for ICRG, or WJP. Specifically, if we look at the partial correlation coefficient of CPI and WGI controlling for the effect of ICRG or WJP, we obtain the extremely high coefficients (0.90 and 0.94, respectively). Such results imply that neither ICRG, nor WJP has substantial intermediate effect on the relationship between CPI and WGI, which in turn means that either they capture only some limited number of corruption activities or they measure some other latent variable; both possibilities imply construct invalidity.

In addition, if we consider partial correlation coefficients of CPI and ICRG controlling for WJP (R = 0.72) and CPI and WJP controlling for the effect of ICRG (R = 0.54), then since the latter is significantly smaller than the former, one may say that ICRG measures corruption more comprehensively than WJP. The same pattern holds in the partial correlation coefficients of

WGI and ICRG while fixing the effect of WJP (R = 0.74) and WGI and WJP controlling for ICRG (R = 0.60). Therefore, we can conclude that among all indices WJP presents the poorest estimation of corruption.

In this section, we showed that ICRG and WJP are the poorest measurements of corruption among the indices we analyzed. In addition, among all indices WJP performed in the way that proves the invalidity of the index. This result may come from the warning that we have discussed above -- acquiring data only from three biggest cities and selection of the non-business respondents answering business-specific questions may have caused bias in the estimations of the level of corruption.

The results on the relatively low construct validity of the ICRG may prove Williams & Siddique (2008) right in their critics on the ICRG; they claimed that PRS group uses too indirect measure of corruption – the longevity of the rule and the way government came to power – that may cause problems in the precision of the measurements of corruption.

Principal Component Analysis

In this section, we use principal component analysis (PCA) in order to see how indices are associated with each other and how do they form the latent variable of corruption. The idea behind this test is as follows: if all indices of corruption measure the same latent variable, then they should form one-dimensional space that itself would construct a synthetic variable representing the latent variable -- corruption. If indices measure different aspects of corruption or measure not only the level of corruption, but, say, some other aspect of quality of government, then they are likely to form at least two-dimensional space, where every space represents a distinct and independent synthetic latent variable. The latter case implies invalidity of the indices of corruption.

Since WJP has huge share of missing data, which makes the implementation of the principal component analysis impossible, and since we have concluded that WJP fails the validity test, we exclude WJP from the analysis.

Running PCA on the whole sample results in obtaining one principal component (PC) with the proportion of total variance is ≈ 92, meaning that indices form one dimension of corruption that indicates possible validity of the indices. As we look at the standardized component loadings of the indices, which can be interpreted as correlation with principal component, we can see that ICRG has the lowest correlation coefficient with PC (R = 0.91), which confirms the conclusion about low construct validity of the indicator.

Table 3. Results of Principal Component Analysis

	PCA on all sample		**PCA in lowly corrupted countries**		**PCA in highly corrupted countries**	
	1 PC	2 PC	1 PC	2 PC	1 PC	2 PC
Share of variation	0.92	0.06	0.91	0.07	0.67	0.17
Eigenvalue	3.69	0.23	3.60	0.32	2.58	0.63
Correlation with Principal Component						
CPI	0.98	-0.11	0.98	-0.14	0.88	-0.12
WGI	0.97	-0.15	0.98	-0.04	0.90	-0.06
BI	0.97	-0.13	0.96	-0.25	0.71	-0.43
ICRG	0.91	0.42	0.88	0.48	0.70	0.65

Note: PC stands for principal component.

Implementation of PCA in two groups of countries provides us with interesting results. In highly corrupted group of countries, the first principal component explains 67% of total variation, whereas the second principal component explains – 17%, and the third – 11%. Consequently, the set of four indices of corruption constitutes rather three-dimensional space, than one-dimensional space, which can be the evidence of the insufficient construct validity of the indices of corruption. The biggest correlation with the first principal component have WGI (R(WGI,PC1) = 0.90) and CPI (R(CPI,PC1) = 0.88), whereas the smallest correlation have BI (R(BI,PC) = 0.71) and ICRG (R(ICRG,PC) = 0.70).

Removing ICRG from PCA for countries with abundant corruption results in slightly different estimates: the first principal component explains 78% of total variation and the second principal component explains – 16%. Despite the fact that reducing ICRG from PCA slightly decreased the dimensionality of the data (now we do not have the third dimension), the formation of two-dimensional space from three indicators does not indicate construct validity of indicators.

In lowly corrupted group of countries, the first principal component explains 91% of total variation, whereas the second principal component merely accounts for 7%. For lowly corrupted group of countries, correlation coefficients with principal component are much higher than for highly corrupted group (see Table 3).

Such result shows us that indices have lower performance in the highly corrupted countries indicating that indices fail to capture some corruption activities that may be present only in the highly corrupted countries. This bias may also come from the subjectivity of the experts that take part in the surveys. Summing up, in this section we showed that indices of corruption performed well in the countries with low levels of corruption and poorly in the countries with high level of corruption.

Factor Analysis

In the next step of the analysis, we consider corruption as a latent variable that forms the indices of corruption, and analyze it using the factor analysis.

The classical factor analysis provides a researcher with unbiased results conditional only on the presence of multidimensional normality in the data. However, on Figure 2 we can see that there is significant skewness in the data (CPI, WGI, and ICRG) and two-peak distribution (BI and WJP) that violates the assumption of multidimensional normality of the data. Therefore, in order to obtain correct estimates for standard deviation and $\chi 2$ statistics, we use in the analysis robust maximum likelihood estimation and Satorra-Bentler $\chi 2$ statistics.

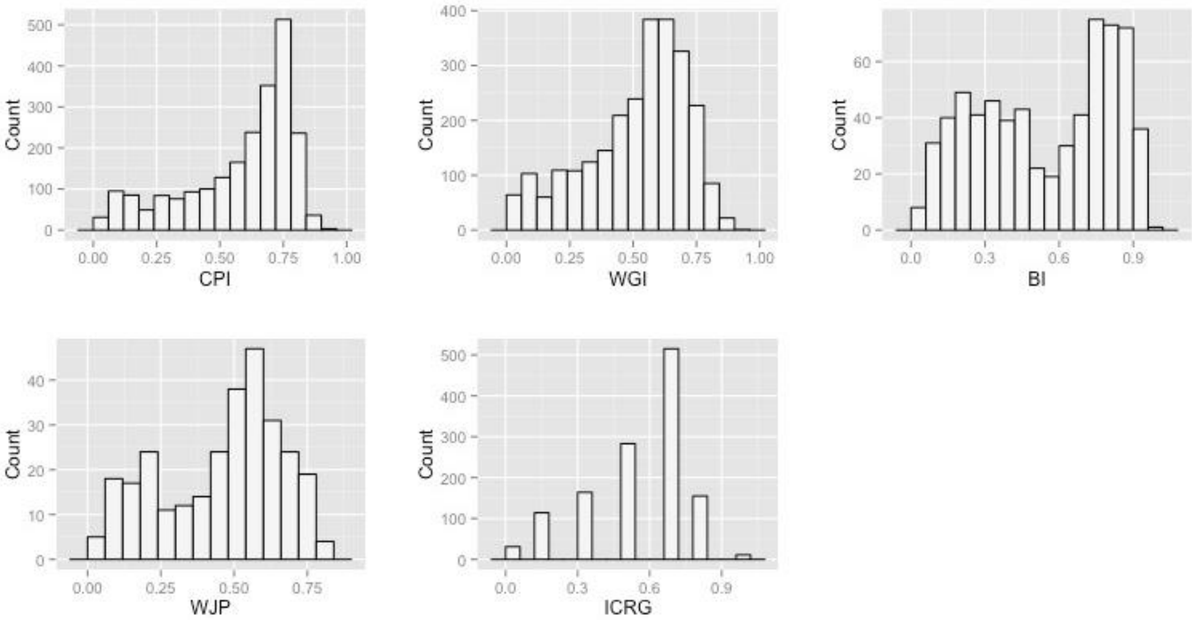

Figure 2. Histograms of indices of corruption

Figure 3 presents a path diagram for factor analysis. Since BI is nested into CPI and WGI, we can assume that there is an interconnection of the stochastic component of BI with CPI and WGI. However, the inclusion of two covariations of stochastic components of BI and CPI, as well as BI and WGI, leads to overindentification of the model. The inclusion only interconnection of the stochastic components of BI and WGI leads to the statistical significance of the difference between the observed and modeled correlation matrices, which means statistical weakness of the model. The insertion of the interconnection of the stochastic components of BI and CPI results in statistical insignificance of the $\chi 2$ test and good model fit. Therefore, we include this interconnection in the model. The inclusion of the additional parameter in the model requires at least four independent variables; therefore, we include ICRG in the model that has previously failed the validity test.

Figure 3 presents standardized loadings of the indices. The highest correlations with the latent variable have CPI and WGI (correlation coefficient with latent factor – 0.99 for both of the indices). BI has slightly lower correlation with latent variable – 0.95 – that may come from the methodology of the index that mostly estimates bribery. ICRG has the smallest correlation with latent variable, and this result corresponds with the previous evidence. Theoretically, BI intends to estimate mostly bribery, whereas ICRG should to estimate a wide range of corruption aspects. The results obtained in CFA show that BI estimates cover the overall corruption more comprehensively than ICRG.

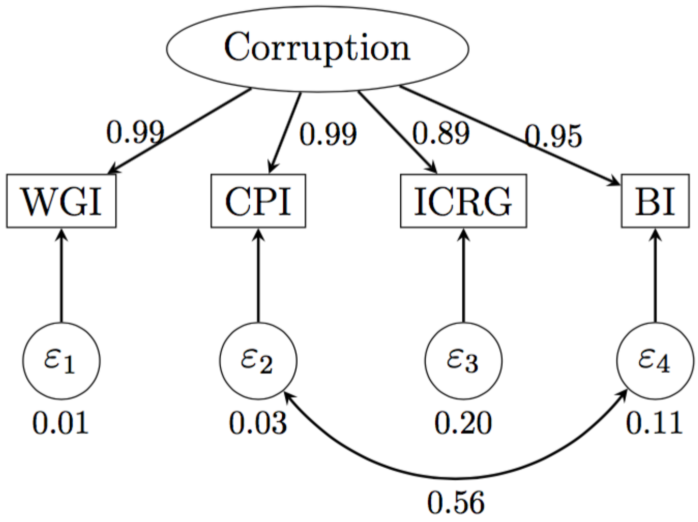

Figure 3. Path diagram for factor analysis.

The substantial advantage of confirmatory factor analysis is that it is able to estimate the reliability of the indices that is the proportion of the indicator's variance left unexplained by the latent factor. In the reliability analysis, the higher reliability of the indicator, the lower the variance of the indicator's error term is. In other words, small variance of the error term indicates high precision of the estimated level of corruption. CPI and WGO have the highest reliability as the variances of their stochastic components are close to zero. Somewhat higher is the reliability of the BI and, finally, ICRG has the lowest reliability. Therefore, we can conclude that ICRG is neither valid, nor reliable indicator of corruption. Whereas, WGI and CPI have the smallest error and, hence, they are the most reliable indices.

Regression

In this section, we analyze the performance of the indices of corruption across countries with low and high level of corruption. If indices are valid, then their relationship should stay the same across different types of countries. If this requirement is violated, then we can state the invalidity of indices.

Figure 4 presents plots of locally weighted scatterplot smoothing (LOWESS) that estimates non-linear relationship between variables. In case of the validity of the corruption indices, we expect that one-unit change in one index will lead to the same change in another index. Substantially, this would imply that two indices provide similar measures of corruption. Moreover, the functional relationship between indicators should be linear and stay the same both in the lowly and highly corrupted group of countries (a diagonal dashed line on a graph of the Figure 4 presents this theoretical relationship). The constant functional relationship between indices in two groups of countries would speak in favor of the absence of biases of the estimation of the level of corruption in either highly corrupted group of countries or lowly corrupted group of countries. In other words, we expect that a pair of constructively valid indices would provide the similar estimation of the level of corruption in all countries from Austria to Zimbabwe.

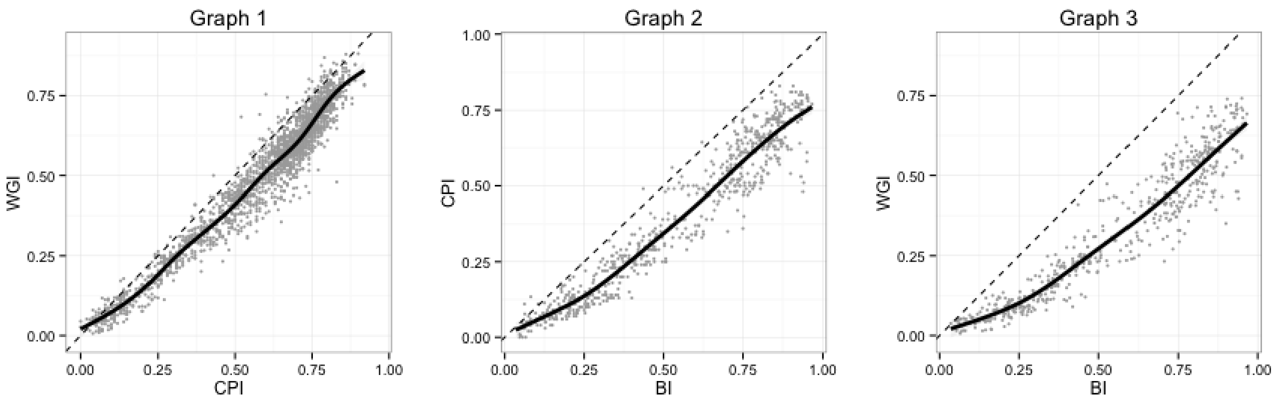

Figure 4. Graphs for LOWESS regression.

Note: Black thin line represents estimates LOWESS regression line, dashed line represents ideal theoretical relationship between the two indices.

On Figure 4 Graph 1 we can see that relationship of CPI and WGI is very close to the theoretical one (the dashed line on the graph). Still, CPI slightly understates the estimates of corruption compared to WGI. In addition, we can see that the distance from the theoretical relationship and estimated one is slightly different across different levels of corruption. When corruption level is close to zero the distance from the theoretical and estimated lines converges to zero and the level of corruption is smaller than 0.20 this distance somewhat increases, but still stays close to zero. On the contrary, once the level of corruption increases, the distance between theoretical and estimated line are evidently non-zero. However, it worth mentioning that in the latter case differences are not dramatic enough to claim that indices are constructively invalid, but significant enough to be cautious about their performance across different groups of countries.

Graph 2 and 3 presents that CPI and especially WGI understate the estimates of corruption compared to BI. In case of WGI, this underestimation is slightly bigger and it increases in the group of highly corrupted countries. From these graphs we can see even more sizeable differences in the distance between theoretical and empirical relationship of indices across different corruption levels. For example, on Graph 3 we can see that only when the level of corruption is very close to zero, the difference between two lines is small, but once we move to the highly corrupted groups of countries this difference increases up to 0.3 units on the 0-1 scale. These results may speak in favor poor performance of BI in the measurement of the level of corruption across different countries. We also can see that the differences between theoretical and empirical relationships across different corruption levels are bigger in case of WGI, implying that compared to CPI, WGI performs poorer.

In the next step of the regression analysis, we estimate regression with interaction term. If indices are constructively valid, then the functional relationship and the slope of the regression line should stay the same across groups with low and high level of corruption.

To test the difference in the functional relationship of the indices across different types of countries, we estimate models where the dependent variable is a corruption index. In this model, the independent variables are another index of corruption and a dummy variable that takes value 1 for the groups of countries with the high level of corruption, and 0 – otherwise. Also we include an interaction term between the index of corruption and the dummy variable in the model to see whether the effect of the independent variable differs across highly and lowly corrupted countries. This would allow us to make conclusions about the validity of the indices across different groups of countries, i.e. the insignificance of the interaction term will be an evidence for the construct validity, since, theoretically, indices should perform equally across different groups of countries. In addition, we intentionally do not add any additional variables as GDP per capita, democracy index, or inflation rate in our regression models to see how these variables affect the corruption indices, since such a model with these variables would suffer from huge endogeneity problems that make any regression coefficient useless and not interpretable.

Table 4. Regression results

Independent variables	Dependent variables		
	CPI	CPI	WGI
	(1)	(2)	(3)
WGI	1.16 (<0.01)		
WGI*HighlyCorrupted	-0.52 (<0.01)		
BI		0.73 (<0.01)	0.57 (<0.01)
BI* HighlyCorrupted		-0.03 (0.60)	0.19 (0.50)
HighlyCorrupted	0.28 (<0.01)	0.14 (0.08)	-0.02 (0.29)
R^2	0.96	0.95	0.93
N	817	259	259

Note: p-values in parantheses.

The results of the first regression model (see Table 4) show that for countries with the low level of corruption one-unit change of WGI leads to the 1.16 change of CPI, whereas in countries with the high level of corruption one-unit change of WGI results in 0.64 change of CPI. This result implies that these indices have different functional relationship across different groups of countries. This speaks in favor of poor performance of these indices in a highly corrupted groups of countries, hence, are not recommended for use in a research that involves highly corrupted groups of countries.

The difference in functional relationship is statistically insignificant for the models (2) and (3) in Table 4 that can be explained by the fact that both World Bank and Transparency International use BI as one of the sources of information. We present graphical representation of the result in Figure 5.

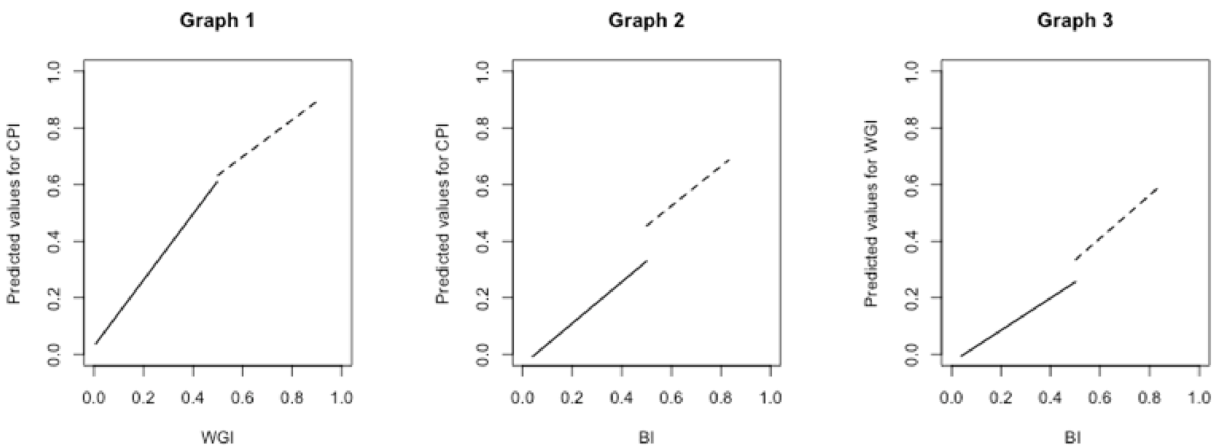

Figure 5. Graph of predicted values.
Note: solid line represents regression line in the lowly corrupted group of countries, dashed line represents regression line in the highly corrupted group of countries.

Discussions

Nowadays indices of corruption are widely used both in the research and in everyday political communication. However, the abundance of such indices and their different measurement of corruption raises a question as to which index can we use with confidence in the research, analytic reports and political communication? Which index better represents the level of corruption across different countries? Is their use in the research at all justified or should we better find some other ways to analyze or control for this aspect of quality of government? In this research, we partially answered these questions.

We analyzed five the most widely used indices of corruption – The Corruption Perceptions Index (Transparency International), The Control of Corruption Index (World Government Indicators), The Bribing and Corruption Index (The International Institute for Management Development), The Corruption Index (International Country Risk Guide), and The Rule of Law: Absence of Corruption (World Justice Project).

In order to estimate the appropriateness of the indices' estimations of corruption, we used the concept of construct validity, where the synthetic measurement of some latent concept is only valid if it correlates with other synthetic measurements of the latent concept. To scrutinize the validity of the indices of corruption we used partial correlation analysis to capture the intermediate effect of one index on the correlation of the other two indices. The idea behind this is as follows: if an index is a valid and sufficient representation of corruption, then controlling for its effect will leave no significant variation in the correlation of the other two indices. Then we ran principal component analysis to see how indices are associated with each other and how many dimensions they form. Ideally, if all indices of corruption are valid, then they should form one dimension that would then construct the synthetic variable representing corruption. Next, we conducted factor analysis to measure the reliability of the indices. Reliability of the index means that the variance of its stochastic component is low enough to make precise conclusions about the level of corruption. In this case, we treat the level of corruption as a latent variable that explains some part of the variation in the index of corruption. The bigger share of the variance of the latent variable, and the smaller its error, the more reliable is the index of corruption. Finally, we ran several regressions with the interaction terms with dummy on the highly corrupted group to check the performance of the indices of corruption across countries with low and high level of corruption. If indices are valid, than their association should stay the same across different types of countries.

In this paper, we showed that Corruption Component of the International Country Risk Guide and Absence of Corruption Index of the World Justice Project are the least valid and reliable indices of corruption among those five indices that we have analyzed. Corruption Perceptions Index and Control of Corruption proved to be the most valid and reliable indicators. However, we also showed that almost all indices have good performance on the lowly corrupted group of countries, whereas their performance on the highly corrupted groups was not acceptable. We showed that the most widely used indices in the academic research, Corruption Perceptions Index and Control of Corruption Index, tend to estimate the level of corruption in highly corrupted groups of countries in different ways. For instance, compared to Corruption Perceptions Index, Control of Corruption Index by World Governance Indicators tends to systematically underestimate the level of corruption. If two indices were perfectly constructively valid, we would never found such statistically significant differences. Unfortunately, on the basis of the test presented in this paper, we cannot say which of these two indices performs better, therefore we can only caution against their use in the countries with abundant corruption.

This conclusion implies that indices of corruption should be reconsidered in a way to make them better estimate the level of corruption in the countries where corruption is ample. Further research in this area may be aimed at distinguishing the possible ways of the amelioration of the measurements of corruption in highly corrupted countries. Perhaps one may test whether one can adjust different surveys to lowly corrupted groups of countries in such a way that they would provide deeper understanding of corruption processes in these countries.

The methodology of the validity evaluation presented in this paper is universal in a sense that it can be used for other social science concepts, such as the level of democracy, rule of law, and other concepts of the quality of government.

References:

1. Botero, J. C., & Ponce, A. (2011). Measuring the rule of law. *Available at SSRN 1966257*.
2. Corruption Prevention and Combating Bureau. (2006) Progress and Results in Preventing and Combating Corruption in Latvia. Riga, Latvia. Retrieved from http://www.knab.gov.lv/uploads/eng/periodic_update_june_2006.pdf
3. Donchev D., & Ujhelyi G. (2014) What Do Corruption Indices Measure? *Economics and Politics*. 26(2), 309-331.
4. Galtung F. (2005) Measuring the immeasurable: boundaries and functions of (macro) corruption indices. In: Galtung F, Sampford C (eds.) *Measuring corruption*. Ashgate, UK.
5. Cronbach L. J., & Meehl P. E. (1955) Construct validity in psychological tests. *Psycho-logical Bulletin*, *52*, 281 – 302.
6. IMD Competitiveness Yearbook, 2014
7. Kaufmann, D., Kraay, A., & Mastruzzi, M. (2011). The worldwide governance indicators: methodology and analytical issues. *Hague Journal on the Rule of Law*, *3*(2), 220 – 246.
8. Ko, K., & Samajdar, A. (2010) Evaluation of international corruption indexes: Should we believe them or not. *The Social Science Journal*, 47, 508 – 540.
9. Krueger, A. O. (1974) The Political Economy of the Rent-Seeking Society. The *American Economic Review*. 64(3), 291 – 303.
10. Kuran, Timur. (1995). *Private Truths, Public Lies: the Social Consequences of Preference Falsification.* Harvard University Press
11. Nisnevich, Y., & D. Stukal (2012). Validity of Public Corruption Measurement in Research of International Organizations. *Problemniy analiz i gosudarstvenno- upravlencheskoe proyectirovanie,* 5(2), 58 – 71.
12. Philp M. (2006) Corruption Definition and Measurement. In *Measuring Corruption* by Ch. Sampford et. al. Ashgate Publishing.
13. Shleifer, A., & Vishny, R. W. (1993) Corruption. *The Quarterly Journal of Economics*. 108(3), 599 – 617.
14. Transparency International. (2011) Corruption Perceprions Index. Long Methodological Brief. Retrieved from http://www.transparency.org/cpi2011/in_detail .
15. Wilhelm P.G. (2002) International Validation of the Corruption Perceptions Index: Implications for business ethics and entrepreneurship education. *Journal of Business Ethics*, 35(3), 177 – 189.

16. Williams, A., & Siddique, A. (2008). The use (and abuse) of governance indicators in economics: a review. *Economics of Governance*, 9(2), 131 – 175.

Contact details and disclaimer:

Anastasiia Shukhova
University of Mannheim (Mannheim, Germany). Research Assistant at the Department of Quantitative Methods in Social Sciences;
E-mail: ashukhov@mail.uni-mannheim.de