# Measuring the Effects of Open Data on the Level of Corruption

Abstract

The Open Government movement has been building around the world with the primary aim to enhance the availability of public data and increase the government's transparency and accountability. Theories of corruption then suggest that higher levels of transparency are associated with lower levels of corruption. Thus, this paper is focused on the potential effects of open data in reducing the level of corruption. For this purpose, different related open data and corruption indices from two different time periods are examined. In contrast to previous studies, the use of open data indices to verify this relationship is new in research on e-government development. Findings indicate that there is a positive correlation relationship between selected corruption and open data indices. Thus, higher levels of open data availability are associated with lower levels of corruption in the compared countries, especially in the case of Open Data Barometer index. It was also found out that the methodologies behind open data indices may affect the results, because they are still evolving due to the increasing effects of open data in the society.

*Keywords*: open data; corruption; indices; correlation analysis; regression analysis

JEL Classification: D73, H11, L86

## Introduction

In recent years, transparency has become an integral part of a broader Open Government movement in which the government acts as an open system and interacts with its environment [7], [13]. Fighting corruption is a very challenging and difficult. Negative effects on development of gross domestic product, unemployment rate or credibility of the country discouraging foreign investors have been shown as a result of corruption [3], [9], [11]. There can be found a number of anti-corruption initiatives around the world that promote greater transparency and openness with the aim to reduce corruption, increase government accountability and improve the quality of public services [4], [13]. For example, the United Nations’ report on e-government has a goal to substantially reduce corruption and bribery in all their forms by promoting the availability of government data, including open data, on online websites, which then helps develop the justice system [13]. Besides that, making government data open is also considered by many to provide greater returns on public investment, help policy-makers address complex problems, creation of trust in government, improve public policies and the efficiency of public services, etc. [2], [14].

On the other hand, simply publishing open data will not necessarily result in a more open, transparent government. Opening data that have no adequate information quality can result in discussions, confusions, less transparency and even in less trust in the government [2], [7], [14]. Rajshree and Srivastava [11] argue that open government data can help in fighting corruption through novel applications that promote transparency in public services. In similar lines, Mistry and Jalal [9] claim that an important strategy for dismantling corruption can be the providing of easy access to information for all citizens through the use of related initiatives. Therefore, to fully exploit the potential of open data, governments have to follow the basic principles of open data. According to Janssen et al. [2], open data should be defined as non-privacy-restricted and non-confidential data which are produced with public money and are made available without any legal restrictions on their usage or distribution. These data should be accessible in both human- and machine-readable formats that allow data to be combined and utilized in different ways using computer programs [2], [14]. Rajshree and Srivastava [11] also noted steps which different stakeholders must take to promote open data and make their potential a reality.

Máchová and Lněnička [7] then evaluated the impacts of open data enabling factors and generating mechanisms in the economic, educational, environmental, health, politics and legislation, social, and trade and business development. Their results suggest that the biggest impacts of open data can be found in the educational and social development, however, the attention of businesses is still lacking in this area. Zuiderwijk et al. [14] focused on discovering what factors are critical for the publication and use of open data in a particular practical case. Categories of factors that were most critical for open data publication in this initiative referred to legislation, regulation and licenses, sustainability of the open data initiative, and accessibility, interoperability and standards. Success factor categories that are critical for open data use concerned legislation, regulation and licenses, and success stories.

The matter of corruption is very often solved in the broader context of e-government [1], [3], [5], [9], [12]. The literature focused on e-government in relation to corruption suggests that electronic service delivery can reduce corruption by minimizing the interactions with officials, accelerating decisions and reducing human errors [3], [9], [12]. However, to the author’s best knowledge, nobody has yet exploited the effects of open data to support this relationship. This is mainly important if Open Government initiatives are much easier and quicker to implement (publish public datasets through open data portals) than robust e-government infrastructures and services [2], [6], [8], [14].

Therefore, the main aim of this paper is to examine the potential effects of open data on the level of corruption. The main reason for conducting this research study of open data and corruption interrelation is given by the recent development of Open Government and open data movement after 2010 and the increased involvement of citizens in the control of government corruption. To estimate the strength and direction of the relationship between open data and corruption, related indices from two different time periods will be examined.

## Material and Methods

As stated by Grönlund and Flygare [1], different indices should be tested to ensure that the effects of open data are robust across different indices and their rankings. Therefore, established indices will be used in order to measure the relationship between open data and the level of corruption [1], [3], [6]. Furthermore, for changes to be discernable there is a need of a few years’ time span [1], [5], [9]. Finally, an adequate data sample is required [9], [12].

Through the last few years, several different indices focusing on the measurement of open data effects were introduced. These are, e.g., the Open Data Barometer index (ODBI) produced by the World Wide Web Foundation (W3F), Open Knowledge Foundation’s (OKF) Global Open Data Index (GODI), the OURdata (Open, Useful, Reusable Government Data) Index introduced by the Organization for Economic Co-operation and Development (OECD) and the European Public Sector (PSI) Scoreboard (PSIS) measuring the status of open data and PSI re-use throughout the European Union (EU) [6], [7]. While it is important to obtain an adequate data sample, the ODBI and GODI will be used as independent variables. The OURdata Index was published only once yet and the PSIS covers only the EU Member States. The Corruption Perceptions Index (CPI) and also the Corruption Control Index (CCI) were then used as dependent variables. Those two indices are the most widely used measures of corruption and their validity have been tested by several researchers. A review of these studies can be found, e.g., in Shim and Eom [12].

The ODBI aims to uncover the true prevalence and impact of open data initiatives around the world. It analyses global trends, and provides comparative data on 92 countries via an in-depth methodology combining related contextual data, technical assessments and secondary indicators to explore multiple dimensions of open data readiness, implementation and impact. The GODI then assesses the state of open data around 122 places in the world and has been developed to help answer such questions by collecting and presenting information on the state of open data around the world to ignite discussions between citizens and governments. Datasets for both these indices are available for the years from 2013 to 2015 and the values range from 0 to 100. Transparency International has published the CPI since 1995. It currently covers perceptions of public sector corruption in 168 countries on a scale from 0 (highly corrupt) to 100 (very clean). Countries’ scores can be helped by Open Government where the public can hold leaders to account, while a poor score is a sign of prevalent bribery, lack of punishment for corruption and public institutions that don’t respond to citizens’ needs. The World Bank’s CCI is one of the six broad dimensions of governance for 215 countries since 1996. It 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. The CCI takes values between -2.5 (weak) and 2.5 (strong) governance performance.

At first, the relationship between open data and corruption in two different years: 2013 and 2015 will be examined. For this purpose, correlation analysis and a simple linear regression model will be used. Correlations between defined variables will be measured by the value of the Pearson's correlation coefficient. According to Nardo et al. [10], several correlation measures (measures of association) can be used to validate the conformity of the rank methods for the indices, such as Pearson’s correlation coefficient, Spearman’s rank correlation coefficient or Kendall’s rank correlation coefficient, Correlation ratio or Mutual information. In this study, Spearman’s and Kendall’s rank correlation coefficients will be used. Contrary to the Spearman’s coefficient, the Kendall’s coefficient is not affected by how far from each other ranks are but only by whether the ranks between observations are equal or not [10]. Then, the following regression function will be used to verify the relationship between selected indices. The function is based on the least squares method and was previously applied by Knězáčková and Linhartová [3], Lupu and Lazar [5], and Mistry and Jalal [9]. It is defined as illustrated in the equation (1):

 (1)

where the dependent variable *corruption\_index* is represented by the CPI or CCI in both years and the independent variable *open\_data\_index* is represented by the ODBI or GODI. The parameter *α* determines the distance of intersection of the regression line with the y-axis (the value of the regression function for x = 0). The parameter *β* is called the regression coefficient and shows the variation of the dependent variable value when the value of the independent variable changes. The symbol *ε* is the residual variance, which is a graphical representation of the distance of points from the regression line [3], [9].

The analysis is performed on the sample of 92 countries, regardless of their geographic location or political regime, in order to determine if open data have improved the level of corruption. After the establishment of regression models for the selected years, a relationship between the change in open data and corruption in this period will be explored. Data collection was made through open sources. All calculations and graphics are done in Statistica 10.

## Results and Discussion

At first, relevant statistical indicators for model variables are presented in Table 1. In the case of mean value, there can be seen an increase between 2013 and 2015, only the GODI has decreased because of the changes in the methodology. The higher mean compared to the median shows that the distribution of values for all these indices is skewed to the left, i.e., there are more countries with lower values than higher ones.

**Table 1. Descriptive statistics (Source: Author)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **GODI 2013** | **GODI 2015** | **ODBI 2013** | **ODBI 2015** | **CPI 2013** | **CPI 2015** | **CCI 2013** | **CCI 2015** |
| Mean | 47.85 | 39.76 | 32.47 | 34.09 | 50.26 | 50.89 | 0.28 | 0.28 |
| Median | 43.00 | 37.50 | 27.58 | 28.73 | 45.50 | 46.00 | 0.02 | 0.09 |
| Std. deviation | 16.80 | 17.21 | 22.67 | 24.12 | 20.28 | 20.70 | 1.08 | 1.07 |
| Minimum | 20.00 | 3.00 | 0.00 | 1.43 | 18.00 | 17.00 | -1.27 | -1.55 |
| Maximum | 94.00 | 78.00 | 100.00 | 100.00 | 91.00 | 91.00 | 2.39 | 2.27 |

Then, the relationship between selected indices in two time periods (2013 and 2015) is examined. Here the null hypothesis defines that the compared variables are not in correlative relationship. Verification of this hypothesis is based on the subsequent comparison of the level of significance with a p-value. In Table 2 is shown a matrix of Pearson’s correlation coefficients on the significance level 0.05, giving a value between +1 and −1, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation. Thus, a positive correlation relationship was found among the variables in both years. This led to the rejection of the null hypothesis. It may be suggested that there is a relationship between the level of corruption and the availability of open data as represented by selected indices in the compared countries. The ODBI has also stronger relationship with both CPI and CCI than GODI. Thus, it furthermore suggests that the ODBI is a better predictor of the level of corruption than the GODI.

Table 3 and Table 4 with Spearman rank order correlations and Kendall tau correlations then confirm these results. All the correlations in both tables are significant at the 0.05 level. Indices focused on the level of corruption rank countries similar to each other while indices focused on open data are slightly different. It may be affected by the first release of these indices, because they were both introduced in 2013 and only three rankings were published yet. Also, the methodology of these indices is still not fixed in time and may change to cover more related attributes. For example, since 2015, four new thematic datasets were added to the GODI. Thus, rank order correlations between GODI 2013 and GODI 2015 are 0.767 and 0.593 while rank order correlations between ODBI 2013 and ODBI 2015 are 0.923 and 0.764.

**Table 2. The matrix of Pearson’s correlation coefficients (Source: Author)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **GODI 2013** | **GODI 2015** | **ODBI 2013** | **ODBI 2015** | **CPI 2013** | **CPI 2015** | **CCI 2013** | **CCI 2015** |
| GODI 2013 | 1.000 |  |  |  |  |  |  |  |
| GODI 2015 | 0.790 | 1.000 |  |  |  |  |  |  |
| ODBI 2013 | 0.851 | 0.792 | 1.000 |  |  |  |  |  |
| ODBI 2015 | 0.796 | 0.871 | 0.905 | 1.000 |  |  |  |  |
| CPI 2013 | 0.679 | 0.648 | 0.707 | 0.735 | 1.000 |  |  |  |
| CPI 2015 | 0.680 | 0.649 | 0.705 | 0.735 | 0.992 | 1.000 |  |  |
| CCI 2013 | 0.676 | 0.666 | 0.712 | 0.752 | 0.996 | 0.989 | 1.000 |  |
| CCI 2015 | 0.645 | 0.622 | 0.686 | 0.704 | 0.992 | 0.991 | 0.991 | 1.000 |

**Table 3. Spearman rank order correlations (Source: Author)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **GODI 2013** | **GODI 2015** | **ODBI 2013** | **ODBI 2015** | **CPI 2013** | **CPI 2015** | **CCI 2013** | **CCI 2015** |
| GODI 2013 | 1.000 |  |  |  |  |  |  |  |
| GODI 2015 | 0.767 | 1.000 |  |  |  |  |  |  |
| ODBI 2013 | 0.855 | 0.795 | 1.000 |  |  |  |  |  |
| ODBI 2015 | 0.817 | 0.830 | 0.923 | 1.000 |  |  |  |  |
| CPI 2013 | 0.603 | 0.520 | 0.708 | 0.722 | 1.000 |  |  |  |
| CPI 2015 | 0.606 | 0.548 | 0.699 | 0.726 | 0.985 | 1.000 |  |  |
| CCI 2013 | 0.579 | 0.526 | 0.728 | 0.734 | 0.989 | 0.982 | 1.000 |  |
| CCI 2015 | 0.546 | 0.485 | 0.706 | 0.714 | 0.981 | 0.984 | 0.986 | 1.000 |

**Table 4. Kendall tau correlations (Source: Author)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **GODI 2013** | **GODI 2015** | **ODBI 2013** | **ODBI 2015** | **CPI 2013** | **CPI 2015** | **CCI 2013** | **CCI 2015** |
| GODI 2013 | 1.000 |  |  |  |  |  |  |  |
| GODI 2015 | 0.593 | 1.000 |  |  |  |  |  |  |
| ODBI 2013 | 0.681 | 0.612 | 1.000 |  |  |  |  |  |
| ODBI 2015 | 0.621 | 0.634 | 0.764 | 1.000 |  |  |  |  |
| CPI 2013 | 0.449 | 0.364 | 0.523 | 0.541 | 1.000 |  |  |  |
| CPI 2015 | 0.452 | 0.392 | 0.518 | 0.547 | 0.913 | 1.000 |  |  |
| CCI 2013 | 0.423 | 0.368 | 0.548 | 0.545 | 0.925 | 0.895 | 1.000 |  |
| CCI 2015 | 0.394 | 0.332 | 0.519 | 0.528 | 0.900 | 0.898 | 0.903 | 1.000 |

As presented above, the ODBI has stronger relationship with both CPI and CCI than GODI, hence, this relationship will be investigated more thoroughly. Figure 1 shows a simple linear regression model. There is the ODBI on the x-axis and the CPI (CCI) on the y-axis for both years. This positive relationship is then illustrated in the scatterplot, which is explained by the linear regression line. There can be seen better results in 2015, especially in the improvement of data openness on the national level. Figure 2 shows the same scatterplot with the CCI as dependent variable.

**Figure 1. The relationship between the ODBI and CPI in 2013 and 2015 (Source: Author)**



**Figure 2. The relationship between the ODBI and CCI in 2013 and 2015 (Source: Author)**



As mentioned above, higher values of ODBI indicate bigger effects of open data initiatives and projects, and higher CPI or CCI values indicate less corruption in the country, i.e., it is perceived to be very clean with strong governance performance and a stable political and economic situation. Focusing on the best performing countries, Denmark, New Zealand, Finland, Norway and Netherlands are among the best performers. Therefore, these countries may serve as a source of practical guidance and policy recommendations for the governing and successful implementation of related initiatives. More detailed view on these data is then given in Table 5. It is focused on the relationship between the changes in open data and corruption in the period of 2013 and 2015. The countries are ordered according to their ranking in 2015 for each index. In the case of the GODI, the change is negative due to the new methodology behind this index. A level of data openness represented by the ODBI has increased in this period by 12.6%. The CPI and CCI have also reported an increase in the mean value, i.e., that the anti-corruption activities are supported by more countries than in 2013. Another finding is that the changes observed in the best performing countries are generally similar across the CPI and CCI while the results for the GODI and ODBI are different, reflecting various approaches in opening up government data.

Furthermore, as this empirical relationship is confirmed, a deeper analysis of the Open Government and open data movement should follow. Some of the recent studies suggest that the availability of open data is increased by implementation of an open data portal [8]. It is a web-based system where an interface allows datasets to be uploaded and equipped with high-quality metadata searchable through defined categories, tags or vocabularies.

**Table 5. List of best performing countries focusing on the changes between variables (Source: Author)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rank** | **GODI** | **ODBI** | **CPI** | **CCI** |
| **2015** | **Change 13-15 [%]** | **2015** | **Change 13-15 [%]** | **2015** | **Change 13-15 [%]** | **2015** | **Change 13-15 [%]** |
| 1 | TW | 85.7 | UK | 0.0 | DK | 0.0 | NZ | -2.2 |
| 2 | UK | -19.1 | US | -12.3 | FI | 1.1 | DK | -5.4 |
| 3 | DK | -19.5 | FR | 27.7 | SE | 0.0 | NO | -0.7 |
| 4 | CO | 0.0 | CA | 22.0 | NZ | -3.3 | CH | 1.6 |
| 5 | AU | 1.5 | DK | 6.6 | NO | 1.2 | FI | -1.9 |
| 6 | FI | -6.9 | NZ | 2.7 | NL | 4.8 | SE | -7.2 |
| 7 | UY | 0.0 | NL | 18.0 | CH | 1.2 | SG | -1.7 |
| 8 | US | -26.4 | KR | 31.3 | SG | -1.2 | LU | -1.6 |
| 9 | NL | -13.5 | SE | -19.2 | CA | 2.5 | LI | 15.3 |
| 10 | FR | 6.8 | AU | 0.5 | LU | 1.3 | NL | -6.1 |
| All countries covered | -2.1 |  | 12.6 |  | 2.0 |  | 26.5 |

The first limitation of this research should be the composition of indices, especially when corruption is influenced by many factors [1], [4]. Another limitation is that the available data for open data indices are still quite recent and it may be questioned if the results are sufficiently reliable for the intended purpose. However, this study may serve as a basis for further research, especially in the context of the proposed methodology for measuring open data effects on the level of corruption. Further, more consistent results may be achieved by dividing countries into groups according to their population, income level, unemployment rate, etc., as suggested, for example, by the United Nations’ report on e-government [13]. Therefore, future research will be focused on various factors which may affect the level of corruption in the context of open data.

## Conclusion

This paper has contributed to the discussion of the potential effects of open data in reducing corruption. Results presented in this paper suggest that there is a positive correlation relationship between selected corruption and open data indices. Thus, higher levels of open data availability are associated with lower levels of corruption in the compared countries, especially in the case of ODBI. Open data may dramatically reduce the time and money citizens need to invest to understand what government is doing and to hold it to account. For this purpose, open data portals are deployed by different levels of government as an important element of most open government initiatives. When coupled with emerging technologies such as social media, it is possible to foster collaboration, engagement, and particularly open data reuse.

However, the ability of open data to potentially decrease the level of corruption is limited to the willingness of government and its institutions to recognize the importance of the Open Government and open data movement and adopt proactive initiatives that promote greater transparency, openness, participation and collaboration of involved stakeholders with the aim to reduce corruption. Some of the recommendations include to review and recalibrate information and data policies, particularly security and privacy, data discoverability, access and reuse; build open data portal and data infrastructure; use data and metadata standards and licenses; and foster research and data communities. Furthermore, it was also found out that the methodology behind open data indices may affect the results, because it is still evolving due to the increasing effects of open data in the society. Therefore, this topic requires continuous research.

Finally, governments have also an opportunity to collaborate more with businesses and citizens in developing enhanced services, and to make effective use of new technologies such as cloud computing, Internet of Things (IoT), mobile internet and big data in the context of greater transparency and accountability.

## References

1. GRÖNLUND, Å., FLYGARE, A. M. 2011. The Effect of eGovernment on Corruption: Measuring Robustness of Indexes. In Andersen, K. N, et al. (Eds): *Electronic Government and the Information Systems Perspective*. Berlin Heidelberg: Springer, pp. 235-248.
2. JANSSEN, M., CHARALABIDIS Y., ZUIDERWIJK, A. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*; 29(4): 258-268.
3. KNĚZÁČKOVÁ, R., LINHARTOVÁ, V. 2013. Impact of e-government at the level of corruption. In *The 7th International Days of Statistics and Economics*. Slaný: Melandrium, pp. 621-630.
4. LINDSTEDT, C., NAURIN, D. 2010. Transparency is not Enough: Making Transparency Effective in Reducing Corruption. *International Political Science Review*; 31(3), 301-322.
5. LUPU, D., LAZAR, C. G. 2015. Influence of e-government on the level of corruption in some EU and non-EU States. *Procedia Economics and Finance*; 20, 365-371.
6. MÁCHOVÁ, R., LNĚNIČKA, M. 2015. Reframing E-Government Development Indices with Respect to New Trends in ICT. *Review of Economic Perspectives*; 15(4), 383-411.
7. MÁCHOVÁ, R., LNĚNIČKA, M. 2016. Exploring the Emerging Impacts of Open Data in the Public Sector. In Špalková, D., Matějová, L. (Eds): *Proceedings of the 20th International Conference Current Trends in Public Sector Research*. Brno: Masaryk University, pp. 36-44.
8. MÁCHOVÁ, R., LNĚNIČKA, M. 2017. Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*; 12(1), 21-41.
9. MISTRY, J. J., JALAL, A. 2012. An Empirical Analysis of the Relationship between E-government and Corruption. *The International Journal of Digital Accounting Research*; 12, 145-176.
10. NARDO, M. et al. 2008. *Handbook on Constructing Composite Indicators: Methodology and user Guide*. Paris: OECD Publishing.
11. RAJSHREE, N., SRIVASTAVA, B. 2012. Open Government Data for Tackling Corruption – A Perspective. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, pp. 21-24.
12. SHIM, D. C., EOM, T. H. 2008. E-government and Anti-Corruption: Empirical Analysis of International Data. *International Journal of Public Administration*; 31(3), 298-316.
13. UNITED NATIONS. 2016. *United Nations e-Government Survey 2016: E-Government in Support of Sustainable Development*. New York: UN Publishing Section.
14. ZUIDERWIJK, A., et al. 2015. Open Data Disclosure and Use: Critical Factors From a Case Study. In *Proceedings of the International Conference for E-Democracy and Open Government (CeDEM15)*. Krems: Universität Krems, pp. 197-208.