Alma Mater Studiorum - Università di Bologna
Department of Statistical Sciences ”Paolo Fortunati”

First Cycle Degree/Bachelor in
Statistical Sciences, Curriculum Stats&Maths

Towards Digitalization Of Governments And The Effects Upon Corruption: A Cross-National Longitudinal Analysis

Presented by:
Alessio Crisafulli Carpani
0000878169

Supervisor:
prof. Marco Novelli

Anno Accademico 2021/2022
Session II

Contents

List of Figures

List of Tables

Introduction

Corruption, commonly defined as *"misuse of public position for private benefit"*, is considered to be one of the most significant impediments to economic development, investments and government spending. In fact, commentators and IGOs, like Transparency International, have emphasized that in order to meet the UNSDG 2030 Agenda, corruption must be reduced. SDG 16 includes commitments to fight corruption, increase transparency, tackle illicit financial flows and improve access to information. As such, the goal is critical to the entire 2030 Agenda, because corruption undermines progress on all other SDGs. The United Nations (2019) estimates that corruption, bribery, tax evasion and related illicit financial flows deprive developing countries of around US$1.26 trillion per year. This has recently been reinforced by research showing that corruption reduces global tax revenues by US$1 trillion annually[1].

Particularly with regard of the recovery plans across countries in the world, like the NextGenerationEU[2], to help repair the damages brought by the Covid-19 pandemic. These are once in lifetime chances for countries to emerge stronger and therefore grant a correct management of resources. However, fighting corruption is not an easy task as the determinants of this complex and multifaceted social phenomenon are nested in the roots of several countries. A feasible way to reduce corruption, especially in the public sector, may be to reduce the interactions between officials and the public. This can be achieved by means of digital governments. Most developed E-governments are assumed to reduce corruption by increasing the transparency between the officials and citizens. Fig.1 shows the level of development of the E-Government Development Index across the countries analyzed in this study by 2018.

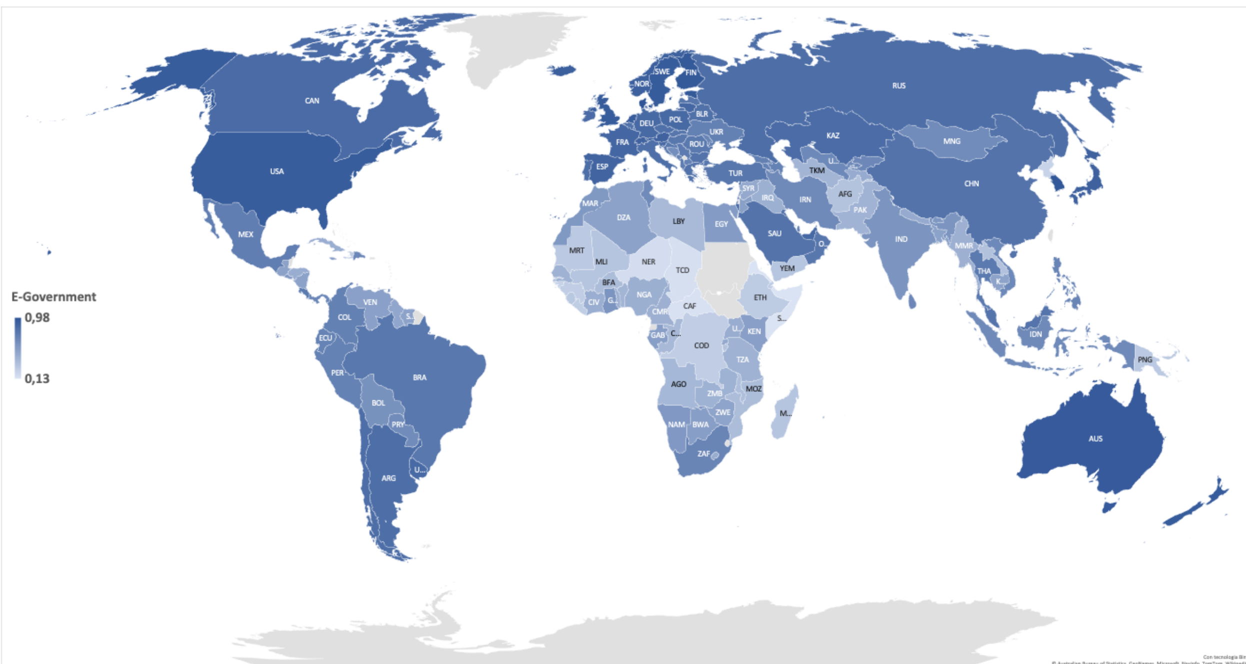


Figure 1: World map of the countries studied by their value of E-Government Development Index

The concept of **e-Government** could be defined as the use of ICT by public institutions to ensure the exchange of information with citizens, private organizations and other public institutions by electronic means, in order to increase the efficiency of their internal functioning and to provide fast, affordable and quality services[1]. The Americas and Asia share almost equal standing in high and middle e-government index levels, and many African countries continue to struggle to improve their e-government standing.

The European countries lead e-government development globally, thanks to the policies implemented by the European Commission[3] to develop the information and communication technologies (ICT) in the field of public administration

[1] SDG 16 is the key to 2030 Agenda
[2] NextGenerationEU
[3] European Commission: eGovernment factsheets anniversary report

(United Nations, 2016)[4]. Besides, one of the pillars in the Digital Agenda for Europe[5] includes “Information and Communication Technology (ICT) to enable benefits for EU Society”, which focus on the improvements that can be made to provide better public services. Basically, this paves the way for more technologically advanced governments, which eventually should foster more efficient and trustworthy governments.

Florini [2], in this view, points out that a better transparency enables citizens to understand a government’s accomplishments. E-government is, therefore, viewed as a positive channel for enhancing trust in government through government accountability and the empowerment of its citizens.

Over the past 10 years, ICTs systems have been developed and adopted for online services, applications, communication and information, and collaboration between governments and non-state stakeholders. Many studies prior to 2010 may not reflect totally this changes. For example, since President Obama’s Memorandum for the Heads of Executive Departments and Agencies in March 2009, that allowed non-state stakeholders to participate more actively in policy decision-making and implementation processes by using ICTs, this transparent open government has attracted an enormous amount of public attention. The implementation of the UN Convention Against Corruption, which has now reached near-universal ratification, providing a comprehensive global framework to fight corruption in alignment with the 2030 Agenda.

One of the studies who focused on one case of e-government initiatives to understand its relation with corruption, was made by Kim et al. in 2009 [3], who studied the OPEN (Online Procedures Enhancement for civil applications) system, recognized by World Bank, UN and OECD, of the Seoul Metropolitan Government, an enhancing administrative system; it is an online system used to disclose administrative procedures (likely to be related to corruption) to citizens in various public service areas. These procedures are displayed in standardized forms in the OPEN system, which guarantees the equity and objectivity of the administrative officer in charge. As a matter of fact, Korea Independent Commission against Corruption (KICAC) confirmed the reduction in corruption.

ICT is already widely used by government bodies, still e-government involves more than just the tools: it involves rethinking organizations and processes, and changing behavior. For example, the procurement system[6] is prone to corruption and inefficiency, due in part to the lack of administrative capacity of the public administration and to weaknesses in the legislative framework. In Italy it is considered a risk area for corruption and organized crime. Estimation reports that the average km of high speed railway track in Italy costs EUR 61 million to build, compared to approximately EUR 10 million/km in France, Spain or Japan. This divergence may serve as an indicator of the scale of mismanagement and irregularities in the procurement process. Since the implementation of the e-procurement systems, the business-to-government purchase and sale of supplies, work, and services (“gare d’appalto”) through the Internet, costs lowered up to €3 billion in Italy. Therefore modern policies can provide important benefits to this matter.

Data

Choosing measures of institutional quality has been simplified by the Quality of Government Institute (QoG)[7], an independent research institute within the Department of Political Science at the University of Gothenburg. The questions addressed by the QoG Institute are how to create and preserve high-quality government institutions and how the quality may influence the socio-economic condition. The dataset established by the QoG Institute is one of the most complete datasets within this field of research and compile data from different reliable and famous sources such as OECD, World Bank, and UN.

The original dataset is drawn from 75 different sources and it consists of approximately 350 variables that deal with topics related to the concepts of Quality of Government [4]. Indexes across 163 countries were selected from the period running from 2012 to 2019 as it is the time series wherein the variables selected do not present missing values; additionally one measure of corruption, the Corruption Perception Index, cannot be compared before 2012, due to a change in its methodology. Moreover, many background studies who have explored this index before this year should be revised[8] for these reasons. (T.I. [2012][5]).

[4]United Nations: E-Government Survey 2016
[5]Shaping Europe’s digital future
[6]Public procurement – Study on administrative capacity in the EU Italy Country Profile
[7]The QoG Institute
[8]Pre-2012 CPI Scores CANNOT Be Compared Across Time

Variables of the Study

The dependent variable of this study is corruption, while regressors includes the E-Government Development Index, the proportion of population using internet and other measures to asses quality of government. Table 1 reports a summary of the statistics of the variables and their definitions.

Table 1: Description and summary statistics of the variables

Variables	Description	Source	Mean (SD)	Median (Range)
Corruption Perception Index	Perceptions of the degree of corruption as seen by business people and country analysis and takes value of 0 if a country is highly corrupt country and a value of 100 indicates a country without corruption	Transparency International	44.02 (19.85)	38.00 (8.00, 92.00)
Bayesian Corruption Indicator	The BCI index values lie between 0 and 100, with an increase in the index corresponding to a raise in the level of corruption	Sherppa Ghent University	47.27 (17.02)	52.14 (6.71, 74.96)
E-Government Index	It is based on a comprehensive assessment of national online services, telecommunication infrastructure and human capital through surveys and ranges between 0 (least developed) and 1(most developed)	United Nations Department of Economic and Social Affairs (UNDESA)	0.57 (0.21)	0.59 (0.01, 0.98)
Government Effectiveness, Estimate	Measures a government's performance and draws on data sources that reflect the perceptions of a diverse group of respondents. It runs from approximately -2.5 to 2.5, with higher values corresponding to better governance	World Bank, WBGI	0.04 (1.01)	−0.08 (−2.45, 2.24)
Rule of Law, Estimate	Includes several indicators which measure the extent to which agents have confidence in and abide by the rules of society. It ranges -2.5,2.5	World Bank, WBGI	−0.02 (1.03)	−0.26 (−2.42, 2.10)
Individuals using the Internet (% of population)	Internet users are individuals who have used the Internet (from any location) in the last 3 months. The Internet can be used via a computer, mobile phone, personal digital assistant etc.	World Bank	47.35 (29.77)	48.11 (0, 100)

The **Bayesian Corruption Indicator** is a composite index of the perceived overall level of corruption. This measure was selected as main dependent variable for two reasons. Firstly, this index is an augmentation version of the *WBGI Control of Corruption*, commonly used in related research, which shows an increase coverage compared to other measures: a 60% to 100% increase relative to the WBGI and the Corruption Perception Index (CPI). In addition, in contrast to the WGI or CPI, the underlying source data are entered without any ex-ante imputations, averaging or other manipulations; this results in an index that truly represents the underlying data, unbiased by any modeling choices of the composer. Moreover, in the BCI a country will have a higher rank if, and only if, it is significantly more corrupt than at least one country with a lower ranking; that is if less than 5% of all draws, the level of corruption of country A is lower than that of B, we can state that B's level of corruption is significantly lower at the 5% significance level. This allows for a more meaningful and less biased comparison of the level of corruption between countries than a simple ranking would allow, unlike the CPI which is very sensitive to the smallest of differences in the actual scores of countries as it takes a simple average of the available sources for each country. The BCI index values lie between 0 and 100, with an increase in the index corresponding to a rise in the level of corruption. Fig.4 shows the changes in BCI for the OECD countries, marking in red and green respectively an increased level of corruption or the opposite.

Transparency International[9], an organization striving for a world free from corruption and the leading global indicator of corruption in the public sector, develops and publishes the **Corruption Perceptions Index** (**CPI**), which relates to perceptions of the degree of corruption as seen by business people and country analysis and takes value of 0 if a country is highly corrupt country and a value of 100 indicates a country without corruption (Fig.2). The CPI, was first developed in 1995 but the methodology used to construct the index was revised by TI in 2012 [5] to ensure that the scores could be

[9] Transparency International

compared between the years thereafter and is the perceived levels of public sector corruption in a country and territory, as determined by 13 independent opinion surveys and expert assessments over 180 countries. This variable will be mainly used to check robustness of the results, as it is preferred the BCI for the reason explained earlier.

E-government refers to the use of information and communication technologies, particularly web-based Internet applications, by government agencies to enhance the access to and delivery of government information and service to citizens, business partners, employees, other agencies, and government entities. The Division of Public Administration and Development Management (DPADM) of the United Nations Department of Economic and Social Affairs (*UNDESA*) develops and publishes the **E-Government Development Index**[10], which is based on a comprehensive assessment of national online services, telecommunication infrastructure, and human capital through surveys and ranges between 0 (least developed) and 1(most developed) over 190 countries (Fig.1). This index, altogether with the **individuals using the Internet** (% of population), is selected as main regressor to assess digital development of countries (Fig.3).

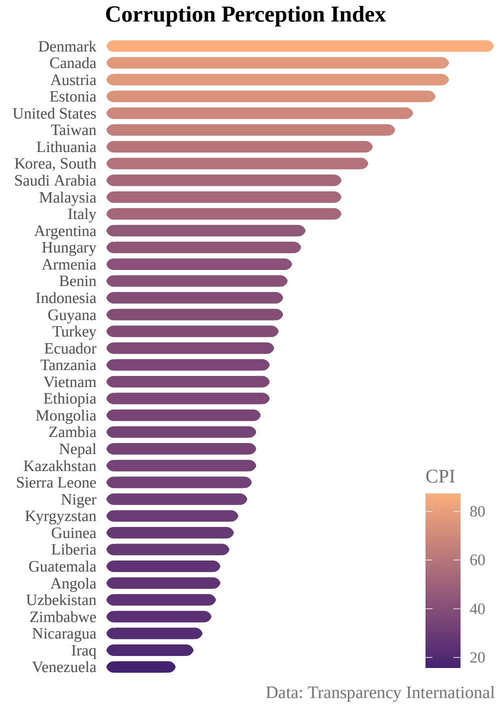


Figure 2: Greatest and lowest quantiles in 2019 for CPI

Government Effectiveness is taken from the World Bank's Worldwide Governance Indicators (WGI). The index runs from approximately -2.5 to 2.5, with higher values corresponding to better governance. The data are based on the one hand on elite perceptions of country experts and on the other on surveys carried out by domestic survey houses as well as cross-national surveys carried out by international organizations and other non-governmental organizations in 212 countries for eight time periods. The World Bank Government Effectiveness Indicator is broad and captures the capacity of a state to implement sound policies by measuring the quality of public services, the quality of the civil service, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.

Another variable provided by the World Bank's Worldwide Governance Indicators, is the estimate of the **Rule of Law**, which includes several indicators which measure the extent to which agents have confidence in and abide by the rules of society. These include perceptions of the incidence of crime, the effectiveness and predictability of the judiciary, and the enforceability of contracts. Together, these indicators measure the success of a society in developing a fair environment where predictable rules for economic and social interactions are respected and the extent to which property rights are protected.

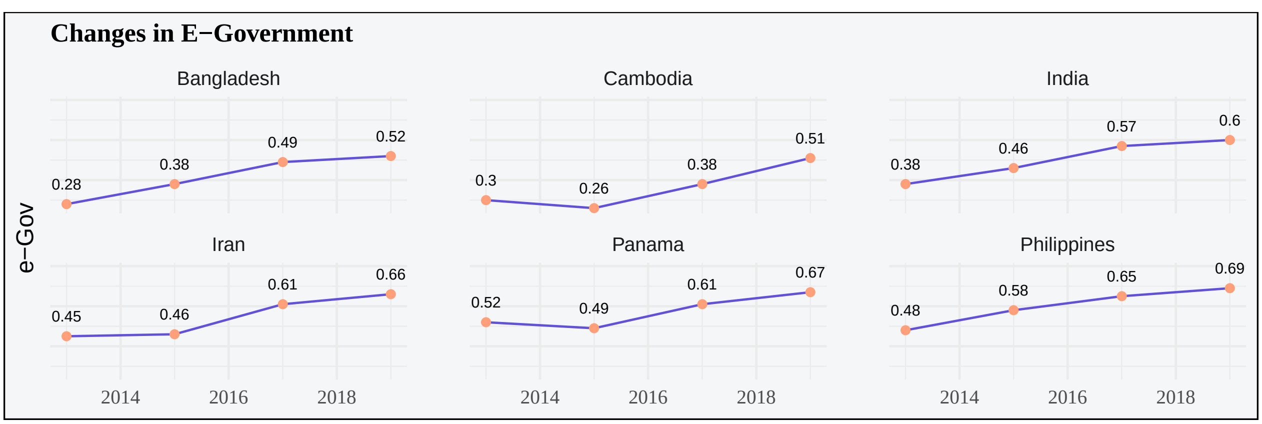
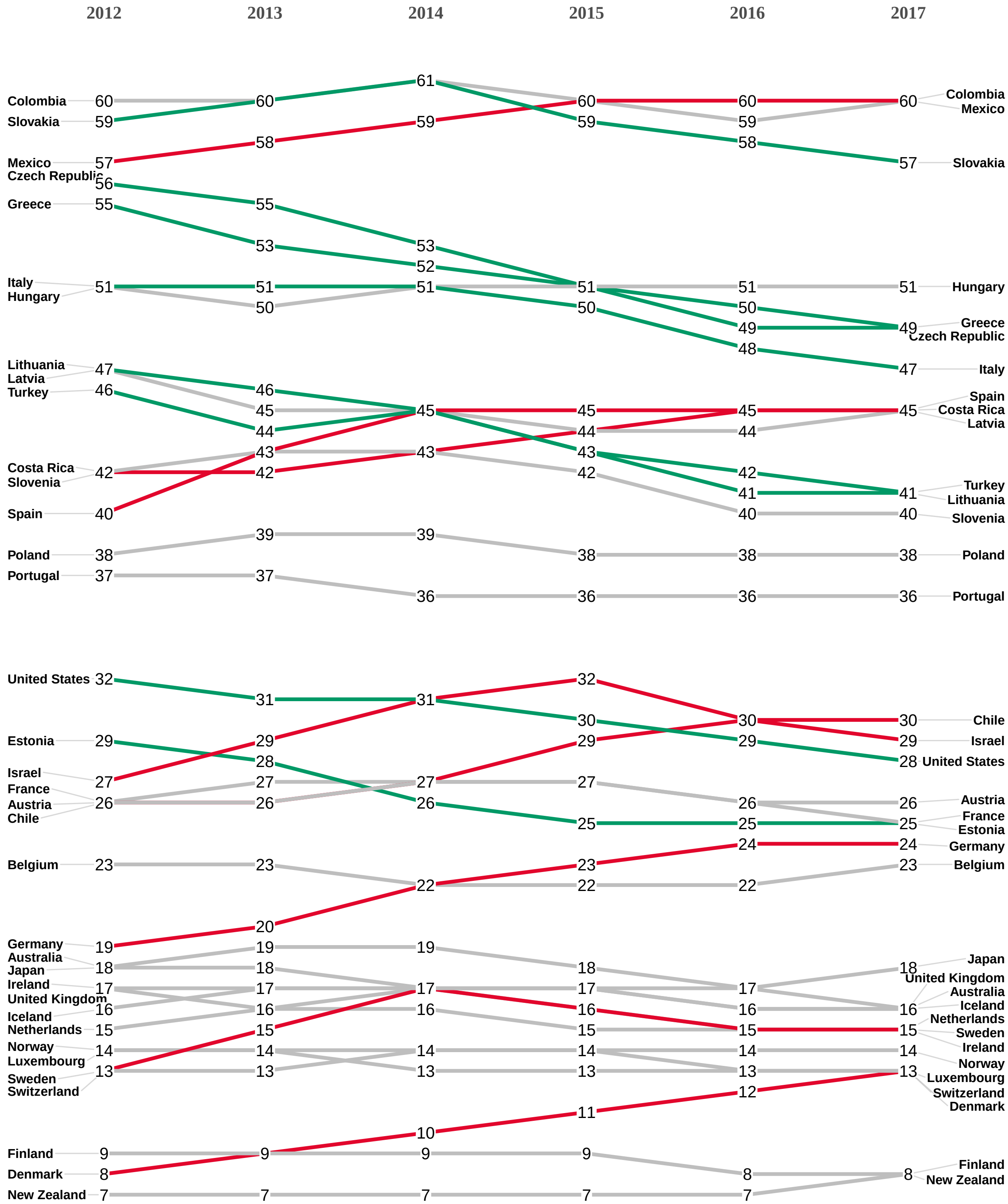


Figure 3: Changes in E-Government Index for selected countries between 2013 and 2019

[10] *DPADM of UNDESA*: E-Government Development Index (EGDI)

Slope Graph of BCI

OECD countries slope changes in the Bayesian Corruption Indicator between 2012 and 2017. Countries who have increased his level of corruption by at least 2 points are marked in red, otherwise in green. Stable position are marked in grey.



Data: S.G. University

Figure 4: Corruption change of OECD countries (2012-2017)

Methodology

Panel Data Regression

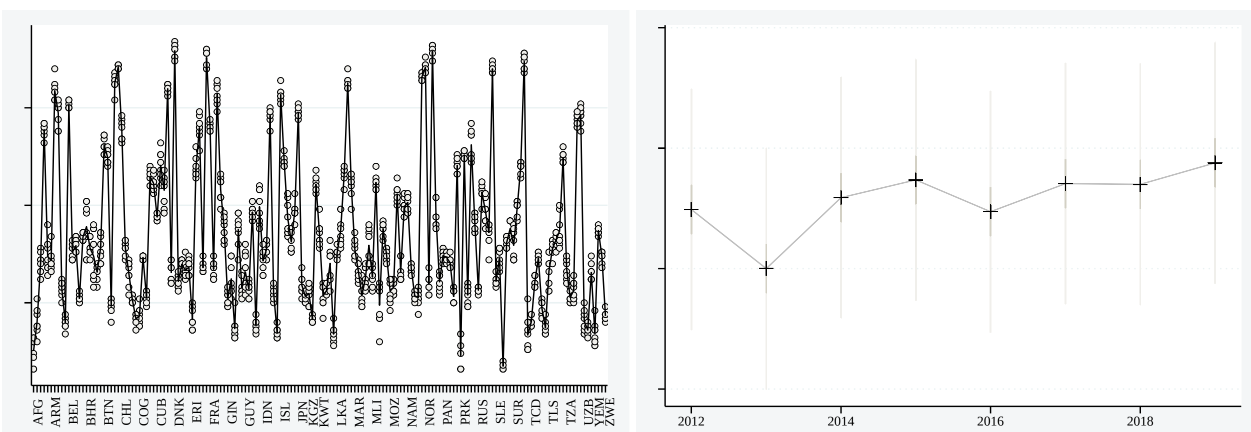
Panel data, also called *longitudinal data* or *cross-sectional time-series data*, is a type of dataset which contains data about multiple variables of interest, observed for a number of entities across several time periods. It contains two types of information, the cross-sectional information which reflects differences between entities and the time series information which reflects changes within entities over time. While ordinary multiple regression techniques can be used on panel data, they may not provide optimal results. Since both cross-sectional dimension and time series dimension of data are involved, there arises the need to examine variation due to individual effects, time effects, or both, which otherwise will lead to biased results. Panel data give more informative data, more variability, less collinearity among the variables, higher degrees of freedom and more efficiency [6].

Sample and considerations

The data used for the analysis is a *balanced micro panel* which means that we have the same number of observations ($N = 1304$) for all the countries (162) during the period between 2012 and 2019

Panel data turns out especially useful if the unobserved heterogeneity (countries' individual characteristics) is assumed time-invariant, otherwise it would affect the *exogeneity* assumptions, that the expected value of disturbances is zero, or disturbances are not correlated with any regressors, and the *homoskedasticity* assumption, that disturbances have the same variance and are not related with one another. The choice of the best model depends also on how one wants to deal with the heterogeneity in the panel, still there are some methods to compute robust standard errors to deal with these issues.

Fig.5a) and Fig.5b) show the fixed effect heterogeneity present across countries and years.



(a) Heterogeineity across countries

(b) Heterogeineity across years

Figure 5: Heterogeneity of Corruption Perception Index

Model

Depending on how we deal with the heterogeneity in the panel, different approaches can arise:

1. *Pooled (Constant Effect) OLS Model* — where the unobserved heterogeneity is considered constant and therefore uncorrelated with the regressors

2. *Random Effect Model* — The differences between individuals are random, drawn from a given distribution with constant parameters, therefore is assumed to be uncorrelated. This method is usually implied when interested in the variance between the individuals of this population, which is not a matter of this study.
3. *Fixed-Effect Model* — The individuals are fixed and the differences between them are not of interest. The heterogeneity is allowed to be correlated with the regressors.

A **Fixed-Effect Model (FEM)** is the best approach for this problem as it is not of interest the differences between countries, but just in the effect of e-government upon corruption. The estimation technique of these models allows to control for these unobservable effects, which are captured by the intercepts, then are demeaned out by the *within transformation* method. This is coherent with this study, as the focus is not concerned about each country specific level.

The FE model is useful in analyzing the impact of variables that change over time. It explores the cause of change within an entity and since time-invariant characteristics are specific to an individual entity and constant within an entity, they cannot cause such a change. Thus fixed effects model controls for time-invariant differences between such entities, such that the coefficient estimators from fixed effects model reflect only the effect of time-varying characteristics and are not biased because of these characteristics as are swiped out during estimation. FE models are preferred for their ability to remove potential confounders and control for omitted variable bias (Wooldrige [6]).

The general equation, for $i\,(=1,...,162)$ countries and $t\,(=1,...,8)$ years, can be written as:

$$Corruption_{it} = \beta\mathbf{X}_{it} + c_i + \lambda_t + \epsilon_{it} \qquad\qquad with: \epsilon_{it} \overset{IID}{\sim} \mathcal{N}(0,\,\sigma_\epsilon^2)$$

where: c_i is the unobserved time invariant individual specific effect that is potentially correlated with the observed regressors but contemporaneously exogenous to the conditional error; that is $E[\epsilon_{it}|c_i] = 0$ for $t = 1, ...T$; λ_t represents unobservable year specific effects such as macrolevel shock in year t; ϵ_{it} is the error term of the model that captures the effect not explained by the regression; α is a scalar constant and β is the regression coefficient vector.

However, before estimating the unknown parameters captured by β, we still need to assume some properties:

(i) $E[\epsilon_{it}|\mathbf{X}_{it}] = 0$ — $\mathbf{X}$ exogenous

(ii) $Var[\epsilon_i|\mathbf{X}_{it}] \neq \sigma^2\mathbf{I}$ — Heteroskedasticity can be allowed

(iii) $E[\epsilon_{it}|\mathbf{X}_{it}, c_i] = 0$ — Strict exogeneity can be imposed

Since an individual specific effect is time invariant and considered a part of the intercept, c_i is allowed to be correlated with other regressors. Here, heterogeneity is assumed to be dealt with countries specific intercepts and the individual effect may possibly be correlated with the regressors, therefore a fixed effect model will better deal with this aspect. Whatever effect the unobserved variables have on the country i at one time t, they will also have the same effect at a later time, thus the supposition is that their effects will be *fixed*. Hausman test to choose between random effects and fixed effects estimation for each regression was run, explained in table 2, and confirming that the fixed-effect model should be preferred.

First off, the data has to be tested for the possible presence of a heteroscedasticity problem with the variance. When conducting the heterogeneity test, there is indeed no constant variance, and the data is heteroscedastic, therefore the least squares estimator is unbiased and consistent but not efficient due to standard errors estimates. In order to control of heteroskedasticity and autocorrelation and make the standard errors robust to these two problems, robust covariance matrix estimators *Driscoll and Kraay* [7] for panel models were computed. Moreover since countries range widely in average across the index, a lot of the variation across all observations in the dataset is accounted by within country-level variation and, additionally, these indexes could lead to a great coefficient of determination as both are constructed upon common characteristics for the countries and therefore one should be aware of correlation issues.

Diagnostic Tests

Several corroborations were applied to the models in order to validate the statistical results. The following diagnostic tests have been implemented through all the models fitted in this study. The table reports briefly the name of the test, its usage and description, reports the test statistics, the *p-value* (computed upon the model specified in table 3(2)) and the consequence in case of rejection of the null hypothesis $H0$ as an example. *Note fully explain the interpretation.*

Table 2: Diagnostic test to assess the assumptions

Test	Description	Statistic	*p-value*	$H1$
Tests for Individual and Time Effects	This test is implemented to check whether there exist any individual effects, time effects or both based on the results of the model.	$F = 164.58$	< 0.01	Significant effects detected, suggesting a two-way FE model.
Hausman Test	*(Fixed vs Random)* Hausman specification test compares a random effect model to its fixed counterpart. If the null hypothesis that the individual effects are uncorrelated with the other regressors is not rejected, a random effect model is favored over its fixed counterpart. It basically tests whether the unique errors c_i are correlated with the regressors, the null hypothesis is they are not. If the null hypothesis is rejected then use fixed effects, if not use random effects.	$\chi^2 = 28.827$	< 0.01	The random effects model is inconsistent, presence of fixed effects detected. Use a fixed-effect model.
Test of Poolability (Chow Test)	Fixed effects are tested by the *Chow test*, If the null hypothesis is rejected the fixed-effect model leads to better results than Pooling OLS. The null hypothesis is that the dataset is poolable so if *p-value* is small enough to reject the null, a fixed-effect model is preferred.	$F = 203.69$	< 0.01	Fixed-effect model is preferred against the pooling one.
Tests for Cross Sectional Dependence (Pesaran's CD Test)	If cross-sectional dependence is present, the consequence is inefficiency of the usual estimators and invalid inference when using the standard covariance matrix (in this case robust covariance matrices are needed).Monte Carlo experiments showed that that the standard Breusch–Pagan LM test performs badly for $N > T$ panels (which is the case), whereas Pesaran's CD test performs well even for small T and large N. Under the null hypothesis of no cross-sectional dependence, $CD \to N(0, 1)$.	$z = -0.43367$	0.664	Strong evidence against $H1$, cross-sectional dependence not detected.
Test of Serial Correlation	Under the null of no serial correlation in the errors, the residuals of a FE model must be negatively serially correlated. The test statistic is obtained by estimating an $AR(1)$ pooling model on the residuals of the within model and then applying a F test. This correlation clearly dies out as T increases, so this kind of AR test is applicable to within model objects only for T sufficiently large. However package `plm` in `r` provides a special test for short FE panels proposed by Wooldridge.	$F = 369.02$	< 0.01	Serial correlation for the idiosyncratic component of the errors detected, use robust standard errors.
Breusch-Pagan Test for Heteroskedasticity	The Breusch-Pagan test fits a linear regression model to the residuals. It tests the null hypothesis that the error variances are equal against the alternative that they are a function of one or more variables and rejects if too much of the variance is explained by the additional explanatory variables.	$\chi^2 = 23.392$	< 0.01	Heteroskedasticity detected, compute robust standard errors.

Note:
The p-value is the probability that the chosen test statistic (which is a measure of how distant are the data and the model prediction) would have been at least as large as its observed value if the null hypothesis were true. Diagnostic test reported very small values of this probability in the F-test of individual and time effect, in the Hausman test, Chow test, Breusch-Pagan test and in the test of serial correlation by Wooldridge. This mean that the data is unusual under the assumptions of pooling OLS models and random-effect models (thus FE model was indeed the best approach), and under heteroskedasticity and serial correlation assumption (for this reason robust standard errors for both heteroskedasticity and autocorrelation were applied). Whereas the Pesaran CD test did report a great value of p-value suggesting that the hypothesis of cross-sectional independence may be compatible with the data, however this value could be inflated because of some other erroneous assumption, therefore standard errors would be robust also for this matter.

Estimation and results

The FE model eliminates the unobserved individual and time effects, which are assumed fixed and that otherwise could not be directly controlled as are not observable, by demeaning the variables using the *within* transformation. Consider:

$\overline{y}_i = \beta\overline{\mathbf{X}}_{it} + c_i + \overline{\lambda} + \overline{\epsilon_i}$

$\overline{y}_t = \beta\overline{\mathbf{X}}_{it} + \overline{c} + \lambda_t + \overline{\epsilon_t}$

$\overline{\overline{y}} = \beta\overline{\overline{\mathbf{X}}} + \overline{c} + \overline{\lambda} + \overline{\overline{\epsilon}}$

Hence it becomes:

$$y_{it} - \overline{y}_i - \overline{y}_t - \overline{\overline{y}} = \beta(\mathbf{X}_{it} - \overline{\mathbf{X}}_i - \overline{\mathbf{X}}_t + \overline{\overline{\mathbf{X}}}) + (c_i - c_i - \overline{c} - \overline{c}) + (\lambda_t - \lambda_t - \overline{\lambda} - \overline{\lambda}) + (\epsilon_{it} - \overline{\epsilon}_i - \overline{\epsilon}_t - \overline{\overline{\epsilon}})$$

$$\implies \ddot{\mathbf{y}} = \beta\ddot{\mathbf{X}} + \epsilon$$

The FE estimator $\hat{\beta}_{FE} = [\ddot{\mathbf{X}}'\ddot{\mathbf{X}}]^{-1}\ddot{\mathbf{X}}\ddot{\mathbf{y}}$ is obtained by an OLS regression of $\ddot{\mathbf{y}}$ on $\ddot{\mathbf{X}}$, and it is unbiased, consistent and efficient. There are costs in the simplicity of this estimation:

- All time-invariant variables for each country are dropped out of the model. This eliminates the between-individuals variability (which may be contaminated by omitted variable bias) and leaves only the the within-subject variability to analyze
- Dependent variables are likely to have smaller variances than in the original specification.
- The manipulation involves the loss of N degrees of freedom (as I am estimating N averages), and too many dummies may aggravate the problem of multicollinearity among the regressors, inflating R^2.

If one is interested in making inference on β, then an estimate of $\sigma^2(\beta)$ is needed. If the error terms ϵ are independent and identically distributed, then the covariance matrix takes the familiar form: $\sigma^2(\hat{\beta}_{FE}) = \hat{\sigma}^2[\ddot{\mathbf{X}}'\ddot{\mathbf{X}}]^{-1}$. The problem at hand is to estimate the covariance matrix of the estimator relaxing the assumptions of serial correlation and/or homoskedasticity (as $Var(\epsilon|X) \neq \sigma^2\mathbf{I}$) without imposing any particular structure to the errors' variance or interdependence. The estimator for the covariance matrix is:

$$\hat{\sigma}^2(\hat{\beta}_{FE}) = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \hat{S}_{SCC} \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$$

Given $i = 1, \ldots, n$ groups, $t = 1, \ldots, T$ period of time, lagged by $l = 1, \ldots, L$, *Driscoll and Kraay* standard errors for the coefficient estimates are then obtained as:

$$\hat{S}_{SCC} = \sum_{t=1}^{T} \mathbf{X}_t^\top \epsilon_t \epsilon_t^\top \mathbf{X}_t + \sum_{l=1}^{L} w_l \left[\sum_{t=1}^{T} \mathbf{X}_t^\top \epsilon_t \epsilon_{t-l}^\top \mathbf{X}_{t-l} + \sum_{t=1}^{T} \left(\mathbf{X}_i^\top \epsilon_t \epsilon_t^\top \mathbf{X}_{t-l}\right)^\top \right]$$

where:

1. The first term, namely $\sum_{t=1}^{T} \mathbf{X}_t^\top \epsilon_t \epsilon_t^\top \mathbf{X}_t$, is the time-clustering heteroskedasticity and cross sectional dependence consistent version proposed by *Arellano* (1987, [8])
2. The second term is the covariance between pairs of observations from any group lagged l periods in time summed with its transpose, i.e.$\sum_{t=1}^{T} \mathbf{X}_t^\top \epsilon_t \epsilon_{t-l}^\top \mathbf{X}_{t-l} + \sum_{t=1}^{T} \left(\mathbf{X}_t^\top \epsilon_t \epsilon_{t-l}^\top \mathbf{X}_{t-l}\right)^\top$. This is a way to control for the effect of common shocks proposed by *Thompson* (2011 [8]), which is then weighted by the Bartlett distance-decreasing kernel function $w_l = 1 - \frac{l}{L+1}$.

Driscoll and Kraay [7] covariance matrix estimators were applied to all the models accordingly to the diagnostic test output; this method is preferred with moderately-sized panel time series in macroeconomics, as it produces heteroskedasticity consistent standard errors that are robust to very general forms of spatial and temporal dependence.

Moreover, to consolidate the validity of the claim, the same model was applied to another corruption measure, the Corruption Perception Index. In table 3, the estimations are shown with the robust standard errors and confidence intervals. The coefficient of e-government turned out to be statistically significant and the direction of the relationship is negative with the BCI, where a unit decrease means a lower level of corruption, and positive with the CPI where the direction is the opposite, so a unit increase means a lower level of corruption.

Hypothesis 1: E-Government developments across countries are able to reduce corruption perception.

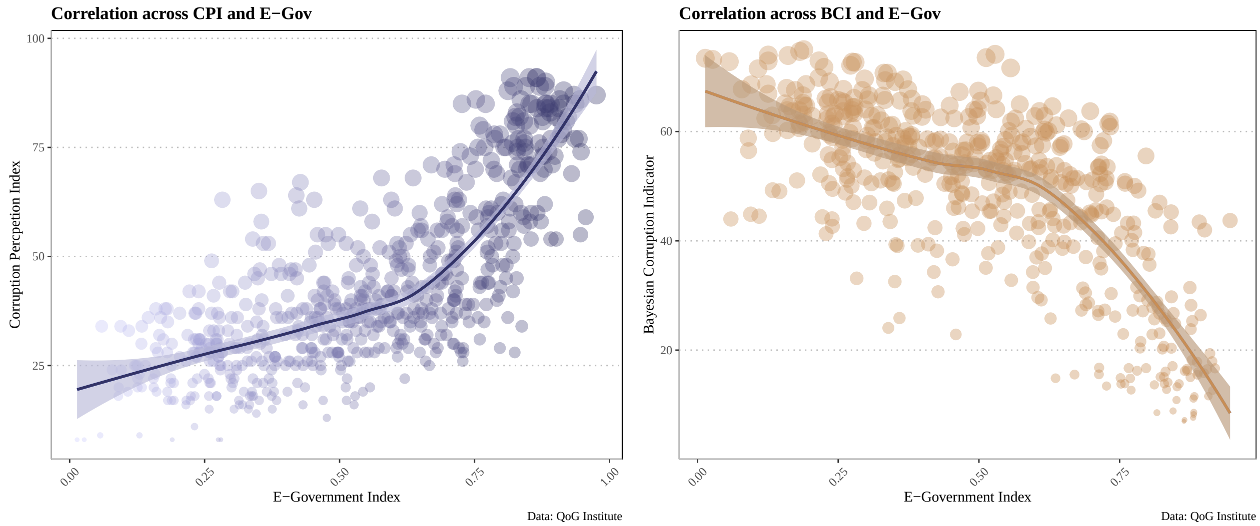
Table 3: The effect of e-government on corruption

	Dependent variable:	
	Bayesian Corruption Indicator	**Corruption Perception Index**
	(1)	(2)
E-Government Index	-3.643^{***}	3.677^{***}
	$(-5.081, -2.205)$	$(1.261, 6.092)$
	t = -4.964	t = 2.983
	p = 0.00001	p = 0.003
Driscoll-Kraay Robust S.E.	(0.734)	(1.232)
Fixed effects?	Yes	Yes
Cross-sectional dependence?	Not present	Not present
Serial Correlation?	Controlled	Controlled
Heteroskedasticiy?	Controlled	Controlled
Spatial Correlation?	Controlled	Controlled
F Statistic	24.641**	6.675*
df	$df1 = 1; df2 = 2$	$df1 = 1; df2 = 2$
R^2	0.512	0.558
Notes:	*p<0.1; **p<0.05; ***p<0.01	
	Confidence Interval for point estimation of 95% level is reported below the estimation, while *Driscoll-Kraay Robust S.E.* are reported after then	

The results from the regression reveals that the expected relationship between e-government development and corruption perceptions have some validity; The t-test for a regression coefficient on the E-government model did detect statistically significant linear dependence of the mean of the dependent variable on the regressor. The sign is concordant with the statement of the hypothesis, it changes between the models according how a smaller level in corruption is measured by the two indexes. The standard errors are small, this indicates that the observed values are close to the fitted regression line and the predictions are more accurate, and so the model. The F statistics, adjusted for robustness, provide strong evidence of goodness of fit among with a reasonable value of R^2, which quantify a good strength of their linear relationship.

For one unit rise in the e-government development index results in a decrease in the mean of the bayesian corruption indicator by 3.64 of a point on average, similarly in magnitude with the CPI. The frequency with which the observed intervals contains the true effect is computed at a 95% level, leaving 5% of chances that the intervals reported would not contain the true estimation, if all the assumptions used were correct.

Fig.6 plots the relationship of E-Government with these two variables.



(a) CPI & EGOV

(b) BCI & EGOV

Figure 6: Correlation of Corruption and E-Government

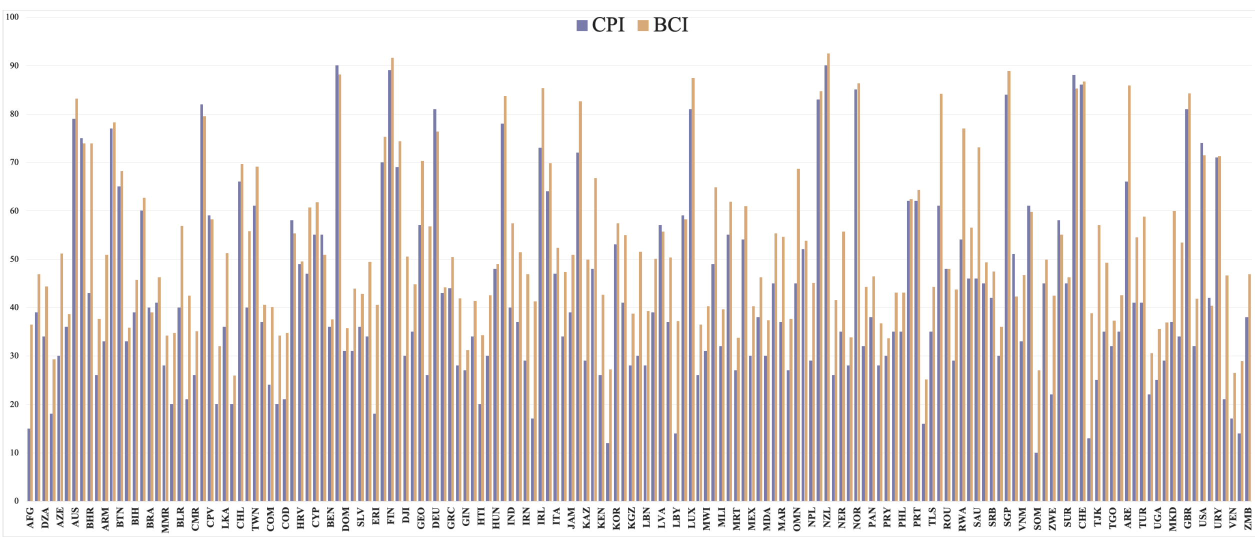


Figure 7: Histogram of the distributions of the two corruption measures

As several other variables could possibly influence corruption, their significance is tested in the added variables analysis to see if their inclusion reduces the significance of the results. Therefore, the aim of next section is to inspect how robust the results are by controlling with other variables, those which previous studies have already found significant with corruption [8]. Also, further investigations on the relationship of corruption with the proportions of ICT users per country are shown, following the path of Heeks [9] who reports that while IT often helps detect and remove corruption, it sometimes has no effect, or creates new opportunities for corruption, like an 'upskilled' corruption.

> *Hypothesis 2: E-Government developments across countries and a greater interaction of the population of Internet may reduce corruption perception.*

The results did not confound the relationship between e-government development and corruption perceptions.

Table 4: The effect of e-government and proportion of IT users on corruption

	Dependent variable:
	Bayesian Corruption Indicator (95% CI)
E-Government Index	-3.404^{***}
	(-4.919,-1.888)
	$t = -4.402$
	$p = 0.00002$
ICT Users (%)	-0.007^{**}
	(-0.013,-0.001)
	$t = -2.308$
	$p = 0.022$
Driscoll and Kraay Robust S.E.	$E - Gov(0.773)$
	$ICT(0.003)$
Fixed effects?	Yes
Cross-sectional dependence?	$Notpresent$
Serial Correlation?	$Controlled$
Heteroskedasticiy?	$Controlled$
Spatial Correlation?	$Controlled$
F Statistic	37.01$**$
R^2	0.526
Adjusted R^2	0.283
p-value:	*p<0.1; **p<0.05; ***p<0.01

The t-test for a regression coefficient on both variables did detect a statistically significant linear dependence of the mean of the dependent variable on the regressors. The sign is concordant with the statement of the hypothesis, it changes between the models according how a smaller level in corruption is measured by the two indexes. The standard errors are small, this indicate more precise estimates. As final criteria, the probability that the goodness of fit F-test statistic would have been at least as large as its observed value, if every model assumption were correct, is reported below the 5% significance level, and therefore quiet reliable. In line with this model, for one unit rise in the e-government development index, the estimated decrease of 3.4 in the mean of the bayesian corruption indicator averagely lowered of 0,20 compared to the single model. In order to answer the question hypothesized, confidence intervals returned useful in addressing the frequency with which the observed intervals contains the true effect. The intervals addressed a value close to 0 in explaining the effect of ICT usage upon corruption, meaning that for the 95% confidence intervals computed upon data from studies would contain a value close of 0 if all the assumptions used to compute the intervals were correct. Also the adjusted R-squared decreased as the term doesn't improve the model fit by a sufficient amount.

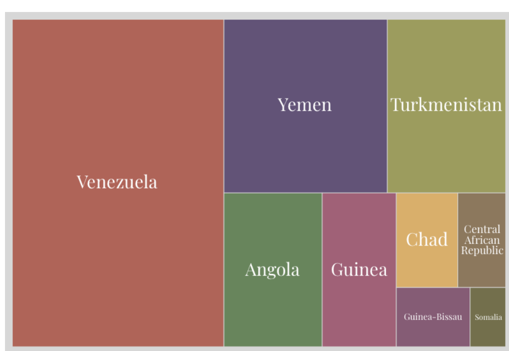


Figure 8: Treemap of most corrupted country by IT usage

Next step consists of controlling for the variables measuring government effectiveness and rule of law from the Word Bank governance indicators [9], also distinguishing by OECD countries or not-OECD.

Hypothesis 3: E-Government across countries with high-income economies (OECD) have a major effect on reducing corruption than developing countries. Controlling for WB indicators of governance.

Table 5: The effect of e-government on BCI and CPI, controlling for government effectiveness and rule of law. Comparing effects worldwide, OECD and not OECD countries

	Dependent variable:					
	Bayesian Corruption Index			**Corruption Perception Index**		
	(Worldwide)	(OECD)	(Not OECD)	(Worldwide)	(OECD)	(Not OECD)
	(1)	(2)	(3)	(4)	(5)	(6)
E-Government Index	-1.742^{***}	-3.448^{***}	-1.513^{***}	2.951^{***}	2.056^{***}	3.379^{***}
	(0.438)	(0.867)	(0.503)	(0.572)	(0.286)	(0.417)
	t = −3.975	t = −3.976	t = −3.008	t = 5.162	t = 7.182	t = 8.103
	p = 0.0001	p = 0.0002	p = 0.003	p = 0.00000	p = 0.000	p = 0.000
WBGI Gov. Effectiveness	-2.918^{***}	-3.792^{***}	-2.845^{***}	3.992^{***}	5.331^{**}	3.817^{***}
	(0.403)	(0.615)	(0.378)	(0.444)	(2.386)	(0.320)
	t = −7.244	t = −6.170	t = −7.520	t = 8.984	t = 2.235	t = 11.922
	p = 0.000	p = 0.00000	p = 0.000	p = 0.000	p = 0.028	p = 0.000
WBGI Rule of Law	-1.302^{***}	0.856	-1.769^{***}	5.796^{***}	3.338^{**}	5.849^{***}
	(0.216)	(0.828)	(0.240)	(0.796)	(1.472)	(0.725)
	t = −6.036	t = 1.034	t = −7.359	t = 7.279	t = 2.268	t = 8.068
	p = 0.000	p = 0.305	p = 0.000	p = 0.000	p = 0.026	p = 0.000
Fixed effects?	Yes	Yes	Yes	Yes	Yes	Yes
Cross-sectional dependence?	Not present	Not present	Not present	Not present	Not present	Not present
Serial Correlation?	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Heteroskedasticiy?	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
Spatial Correlation?	Controlled	Controlled	Controlled	Controlled	Controlled	Controlled
F Statistic	10.82***	2.247*	9.13***	27.71***	3.433**	22.61
R^2	0.754	0.776	0.648	0.916	0.915	0.847
Adjusted R^2	0.627	0.643	0.464	0.887	0.881	0.794
p-value:						*p<0.1; **p<0.05; ***p<0.01

The coefficients of e-government altogether have confirmed again the claim, even after the inclusion of the other variables. The coefficients and models are reliable, except for the F statistic of the model (6). The coefficient is lower compared to the previous models, surely because e-government appears to have a minor effect on reducing corruption than other administrative and political factors do. The same reason could also explain why the coefficient is lower in not-OECD countries; before considering digital developments, these countries need more effective policies and changes. It is important to point out that the two indexes measures shows differences with regard of the model applied to the OECD and not OECD countries, however this could mainly due to the coverage and reliability of this indexes construction, as explained below table1. Still, the coefficients signs confirm the negative relationship with the corruption level (a lower level of corruption is expressed by a negative sign in BCI and a positive sign CPI). The coefficients of government effectiveness and the rule of law as expected sustain the previous findings, that these factors play an important role in mitigating corruption. Keeping all the other variables constants, the estimated coefficients for government effectiveness range (in absolute values) from 2.91 to 3.9 for the worldwide model (1) and (4), meaning that one unit improvement in governance effectiveness leads to a reduction on the mean corruption of between 2.91 and 3.9, and between 1.3 to 5.7 for rule of law, on average. For what concern the e-government, one unit increase in the e-government development index leads to a 1.74 decrease in the mean of the bayesian corruption indicator on average for the worldwide model (and similarly for the not-OECD one), while consists of 3.45 decrease on average for the OECD models. This results stress the idea that more advanced policies in the field of e-government, like e-procurement system, open government, e-participation of citizens etc, could serve as anti-corruption tools in fighting this social phenomena. Moreover most economic developed countries should focus more on the digital transformation, as could have a more significant effect.

Conclusions

This study aimed to estimate the effect of digitalization across governments on corruption. Using a longitudinal panel data, including 162 countries for the period of 2012-2019, two-ways fixed effect model as a data analysis method was employed. The focus on this time series was to reflect the adoptions of new technologies in the field of ICTs by governments in the last 10 years and provides reliable updates in methodology of the indexes implemented, as many backgrounds studies on corruption implemented a prior 2012 version of CPI, unusable for time to time comparison.

Before addressing any important result, several diagnostic tests on the data for the model selected were implemented to rely the most on the consideration made. Accordingly to the test, throughout the study, I have computed robust standard estimators estimated according to *Driscoll and Kraay*. Further, I aimed to investigate if the choice of corruption index matters for the results. The main corruption indicator used in this study is Bayesian Corruption Index, yet an additional corruption measure was included to check the robustness of the results, the Corruption Perception Index.

Findings reflected the the positive impact of e-government in reducing the level corruption, in addition to the fundamental preconditions of a clean government. Even if ICT usage was not founded to be directly efficient in this analysis, successful ICT-enabled transparency systems exist and could be further studied. Also, the OECD and not-OECD models pointed out that for a practical point of view, within least developed countries, an approach which involves both anti-corruption policies and digitalized administrative initiatives may be more effective in reducing corruption than relying just on digital development; whereas OECD countries should invest more and more in digitalization for bureaucratic and governmental matters as its effect could be really positive. However, the research did not take into account educational and economic factors as moderators, thus more complete researches may examine how those factors influence the impacts of e-government on corruption.

Surely, digital governments with high-quality public bureaucracies, consisting of competent public agents, if strengthened with e-government, anti-corruption efforts could more effectively lower the corruption levels of public affairs. For example by establishing a central register of beneficial ownership information with strong verification mechanisms, made publicly available in an open data format. This will help end the abuse of anonymous companies and other legal vehicles that facilitate corruption, preventing the diversion of critical resources needed to fund sustainable development. Transparency can be achieved by providing citizens with more and direct access to information regarding the businesses of the citizens concerned, and supervising the arbitrary behaviors of government officials.

This increased transparency may leads to a decreased level of corruption.

Bibliography

[1] Lidinsky, V. (2008). *EGovernment bezpecne,1*. Praha: Grada Publishing.

[2] Florini, A. (2000). Does the invisible hand need a transparent glove? The politics of transparency.

[3] Kim, S., Kim, H. J. and Lee, H. (2009). An institutional analysis of an e-government system for anti-corruption: The case of open. *Government Information Quarterly* **26** 42–50.

[4] Dahlberg, A. S., Stefan. (2021). *The quality of government basic dataset, version jan21*. University of Gothenburg: The Quality of Government Institute.

[5] Anon. (2012). *Technical methodology note corruption perceptions index*. Transparency International.

[6] M.Wooldridge, J. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

[7] Driscoll, J. C. and Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data.

[8] Millo, G. (2014). *Robust standard error estimators for panel models: A unifying approach*. Munich Personal RePEc Archive.

[9] Bank, W. (2011). Worldwide governance indicators (wgi) project.

[10] Standaert, S. (2015). Divining the level of corruption: A bayesian state space approach.

[11] Linhartová, V. (2016). *The role of e-government in mitigating corruption*.

[12] Kim, C.-K. (2013). *Anti-corruption initiatives and e-government: A cross-national study*.

[13] Park, H. (2003). *Determinants of corruption: A cross-national analysis*. University of Central Florida.

[14] Baltagi, B. M. (2005). *Econometric analysis of panel data, 3rd ed*.

[15] M.Wooldridge, J. (2012). *Introductory econometrics: A modern approach, 5thED*.

[16] Yves Croissant, G. M. (2019). *Panel data econometrics with r*.

[17] Yves Croissant, G. M. (2008). The plm package.

[18] M.Arellano. (1987). Computing robust standard errors for within-groups estimators.

[19] Torres-Reyna, O. (2010). *Getting started in fixed/random effects models using r*. Princeton University.

[20] Per Jonas Partapuoli, E. S. (2019). Corruption kills.

[21] Heeks, R. (1998). Information technology and public sector corruption.

[22] S.Greenl, K., S.Senn. (2016). Statistical tests, p values, confidence intervals, and power: A guide to misinterpretations.