Departments of Economics
NEKH03
Bachelor's Thesis, 15 credits
August 2022

Corruption's Impact on Welfare Spending

A Panel Data Analysis

Author:

Leo Kleja

Supervisor:

Fredrik NG Andersson

Abstract

The existence of corruption poses challenges for governments worldwide attempting to implement certain policies as well as organizing their economy. Not only can widespread corruption threaten the rule of law and legal order but also the trust in and legitimacy of public institutions. The aim of the study was to investigate possible relationships between corruption and public expenditure on healthcare spending, education spending and expenditure on social services. To analyze the effect of corruption on the share of public spending allocated towards these selected welfare policies, a fixed effects panel data regression model was used. Data on corruption from a total of 178 countries was collected through Transparency International's Corruption Perception Index (CPI), an organization working to combat and measure corruption on an international level. The CPI was standardized in 2012, being the logical starting year of this study. In total, the database contained 11 679 observations (including control variables) for the years 2012-2020. The panel data analyses showed that there was a significant negative relationship between corruption and healthcare spending. Similarly, there was a nearly significant negative relationship between corruption and education spending. Thus, the results obtained supports the hypothesis that corruption decrease certain shares of welfare spending as part of public expenditure.

Table of Contents

1. Introduction

The existence of corruption poses challenges for governments worldwide attempting to implement certain policies as well as organizing their economy. Not only can widespread corruption threaten the rule of law and legal order but also the trust in and legitimacy of public institutions. On a high level this is because corruption enables certain individuals to unaccountably break the law or change the law to their benefit. On a lower level, corruption can be part of the daily lives of a countries' citizens when much of local governing and bureaucracy is run, through bribes. Both these levels affect the quality of public services offered by the government in a country. Furthermore, there is evidence in previous research, that will be explored in this essay, that corruption also effects the composition of government expenditure. In other words, the size of share of the government budget that goes to what services. This thesis will investigate if corruption and its consequences affect the share of government expenditure dedicated towards some selected welfare programs, namely healthcare, education, and social services.

The issue of corruption is generally more severe in poorer nations but remains a problem even in developed nations. For example, according to the 2022 Eurobarometer on corruption, (European Commission, 2022) corruption remains a serious concern for the people and entrepreneurs in the EU. This years' edition of the Corruptions Perceptions Index (CPI) from Transparency International also shows that the corruption level in the world have remained at a global standstill during the first years of the covid-19 pandemic (Transparency International 2022a). In fact, the global average score of the CPI has remained the same the last decade (Transparency International 2022b). The definition of corruption used in this thesis is "abuse of entrusted power for private gain". This definition is well established and used by Transparency International (Transparency International 2022c) and the Eurobarometer (European Commission 2020).

There are several economic reasons as to why corruption is detrimental. One way it effects the economy of a nation is through decreasing the efficiency of the government bureaucracy. This is because corrupt government bureaucracies impede the growth and creation of businesses and investment (Mauro 1995). Part of the explanation lies in the difficulty of acquiring patents, securing intellectual property rights, and a slower process for the government to issue permits and licenses in countries with a high corruption level (Ibid).

These difficulties are in turn often caused by so called rent-seeking (Murphy et al. 1993). Rent-seeking is defined by the Britannica (2022) dictionary as the "competition for politically protected transfers of wealth", i.e., different market actors attempting to influence the government to acquire an "economic rent", e.g., patent, subsidy, or monopoly position, often through bribery. Problems in the economy occur when rent-seeking is more profitable than productive methods of profiting, since this lowers productivity (Murphy et al. 1993). Additionally, rent-seeking by government officials in particular may repress innovation, which in turn lessens economic growth (Ibid). In order to explore if this relationship is empirically true, Mauro (1995) used OLS (see section 5.3) and two-stage least squares model on data from multiple countries (see section 2.2) and found that corruption indeed decreases economic growth and deter investment. The implication of this is that reducing corruption should be a priority for any government also trying to grow an economy.

2. Background

2.1 How to Measure Corruption

In order to quantitively analyze the effect that corruption has on government expenditure on welfare policy, a quantitative method on how to measure corruption is needed. There are some researchers arguing that measuring corruption is impossible. (Goel & Nelson 2010, p. 434) (Zaman 2009). Nevertheless, numerous attempts to quantify corruption have been made, starting in 1980 by private companies selling information to investors requesting risk estimates (Hawken & Munck, 2011, p. 1). Other approaches of measuring corruption were evolved by NGOs and IOs after quantitative indicators became increasingly popular (Ibid). This resulted in many new corruption measures being developed, among which the one developed by Transparency International is one of the most well-known (Villarino 2021). Simultaneously, a discussion around "measurement validity" started to take form, adhering to the notion that there are some methods that are better than others (Hawken & Munck, 2009, p. 6). This assumes that there is a science in measuring corruption, although not necessarily a precise one (Villarino 2021). The topic remained to be disputed with De Maria (2008, p. 187) arguing that corruption is nearly impossible to measure due to, among other things, the invisibility of the degree of the problem. At a similar point in time Kaufmann et al. (2007) and Lambsdorf (2007) argued the opposite, i.e measuring corruption is possible with the right method. Furthermore, during the last 15 years corruption measurements and indices have considered increasingly relevant and have been frequently used (Villarino 2021).

The two most widely used corruption measurement metrics today are Transparency Internationals Corruption Perception Index (CPI) and the World Bank's Control of Corruption Indicator (UNDP 2008, p.6). While they both are based on perceptions of corruption gathered in surveys and opinion polls, slightly different approaches are being used. The World Bank surveys business leaders, adding focus on the private sector, while the CPI is solely centered around corruption on the public level (UNDP 2008, p.14).

The index used to measure corruption in this article is the Corruption Perception Index since it is standardized, i.e., uses a very similar method in all countries over time, and is therefore comparable both between nations and between years from 2012 onwards (Transparency International 2021). As previously mentioned, the index reflects the perception of corruption in a country. Perception is used as a proxy for the actual phenomenon, and it is based on expert opinions and surveys from several sources (Ibid). The CPI as of 2021 was constructed from the 13 data sources listed below:

1. African Development Bank Country Policy and Institutional Assessment 2020
2. Bertelsmann Stiftung Sustainable Governance Indicators 2020
3. Bertelsmann Stiftung Transformation Index 2022
4. Economist Intelligence Unit Country Risk Service 2021
5. Freedom House Nations in Transit 2021
6. Global Insight Country Risk Ratings 2020
7. IMD World Competitiveness Center World Competitiveness Yearbook Executive Opinion Survey 2021
8. Political and Economic Risk Consultancy Asian Intelligence 2021
9. The PRS Group International Country Risk Guide 2021
10. World Bank Country Policy and Institutional Assessment 2020
11. World Economic Forum Executive Opinion Survey 2020
12. World Justice Project Rule of Law Index Expert Survey 2020
13. Varieties of Democracy (V-Dem v. 11) 2021

(Transparency International, 2022d)

The following requirements are valid for being a reliable data source: 1) being quantifiable, 2) using a standardized method measuring countries on the same scale, 3) performed by a credible institution, using country experts or businesspeople and 4) the assessment should be repeated at least every two years (Ibid).

There have been several criticisms of using an index like the CPI to measure corruption. Firstly, the meaning of corruption can mean different things to different observers. Therefore, an element of subjectivity is incorporated in the measurement affecting the validity. (Bello y Villarino 2021) Secondly, one should be cautious when looking at changes over a narrow number of years due to time lags between actual changes and the perceptions of corruption in a country (Stephenson, 2015). However, according to the qualitative analysis by Bello y Villarino (2021) CPI is relatively reliable and passes a minimum validity test and can be used to measure corruption, even though certain restrictions exist. He argues that a larger time-period should be used, and more than one country should be included in an analysis using an index such as the CPI.

2.2 Previous Research on the Effect of Corruption on Public Expenditure

One of the earliest studies on the effect of corruption on public expenditure was made by Nice (1986). He analyzed the effect of corruption on different governmental programs in the US during 1976-1980 by using the method of zero-order correlation analysis, i.e., investigating the correlation between two variables, equal to a Pearson correlation. The many variables included in the analysis were: unemployment benefits, AFDC benefits per family, AFDC benefits per recipient, state local education expenditures per capita, state local welfare expenditures per capita, state local highway expenditures per capita, tax effort, education effort, highway effort and welfare effort. Effort was defined as the allocated share of taxes as a percentage of personal income. The definition of corruption used was the number of federal, state, and local officials convicted on federal charges, 1976-1980, per 100,000 public employees.

Nice (1986) discussed three hypotheses: First, that corruption would lead to less funding to programs seeking to benefit the low-income earners. Second, that corruption is a mean to overcome aversion to change. Lastly, that corruption leads to a less effective and increased program costs without increased benefits. He found no evidence for any of the hypotheses. While there where some correlations between corruption and the variables, the relationships were too weak to be significant (Ibid).

The first cross-country study on the influence of corruption on government expenditure was "Corruption and the Composition of Government Expenditure" (Mauro, 1998). In this article Mauro argues that the existence of rents in government will lead to rent-seeking behavior, a hypothesis previously proposed by Kreuger (1974). Therefore, government expenditure in programs that are more prone to rent-seeking may increase when corruption increases. Markets lacking competition are also more vulnerable to corruption through increased bribes. Projects, which costs are difficult to estimate, such as ones involving specialized high technology services are opportune to corruption, including infrastructure projects. The method used by Mauro (1998) was an OLS regression and the variable for corruption used was developed by the International Country Risk Guide, i.e. a series of indices made by the private firm Political Risk Services, Inc and compiled by the University of Maryland. The data was available for 106 countries.

To account for the issue of endogeneity as well as simultaneity, that is only capturing the effect that corruption has on the composition of government expenditure and not the other way around, Mauro (1998) used a number of instrumental variables as proxies for corruption. These were: 1) ethnolinguistic fractionalization index (Taylor & Hudson, 1972), 2) a colony after 1776 dummy variable, 3) an independent after 1945 dummy variable and 4) an oil dummy variable (Barro, 1991). A black-market premium variable (Levine and Renelt 1992) and the ratio of sum of imports and exports to GDP (World Bank Stars database) were also included as an instrumental variable for trade restrictions constituting a source of rents. In a first analysis, univariate regression was used to analyze the relationship between corruption and the different dependent variables. Subsequently, GDP per capita and share of population aged 5-20 was added as control variables in a multivariate regression model. Government spending on education proved to be significantly influenced by corruption at the 1% level and within the 5% (2%) when using robust standard errors. The relationship was negative, i.e. countries with a higher degree of corruption tended to have a lower share of government spending on education. He also found a significant link between corruption and healthcare expenditure. However, when testing with other proxies for corruption the link was only significant at the 10% level.

An issue when dealing with the topic of corruption's effect on public expenditure is the direction of causality (Mauro, 2002). It might be that in countries with a robust safety net it is harder for corruption to persist. It may also be the other way around, i.e., corruption makes it harder to spend money on welfare programs since the money is spent elsewhere. Hence, it is possible that causality operate in both directions (Ibid) and it might be difficult to determine the direction of causality. Mauro (1998) attempted to account for this by using an instrumental variable approach.

The method of instrumental variables used by Mauro has been critiqued by Harvard Professor Matthew Stephenson (2014). He criticizes a previous article of Mauro (1995), where, just as in (Mauro 1998), Mauro uses ethno-linguistic fractionalization as an instrumental variable. Stephenson argue that this variable does not fulfill the criteria for an instrumental variable. A good instrumental variable should only affect the dependent variable through its effect on the explanatory variable. In the case of Mauro (1995), ethno-linguistic fractionalization must affect economic growth only through its effect on corruption. Stephenson argues that this might not be a very credible assumption to make.

Attempting to address the problem of simultaneity Delavallade (2006) used a three-stage least square method developed by (Zellner and Theil, 1962) when analyzing the question; how corruption influences the distribution of public spending in development countries. The method first regresses the endogenous variables, in this case the different shares of public spending and the corruption variable, on all the model's exogenous variables. The result is the predicted values of the endogenous variables, which are used as instrumental variables to the endogenous variables. After the variance-covariance matrix of the residuals is estimated in the second stage, the third stage is a generalized -least square regression using this matrix and the instrumented values of the right-hand endogenous variables.

Delavallade (2006) included several control variables in the analysis, being: 1) the proportion of urban population in the whole population, 2) the dependency ratio of the population, corresponding to the ratio of people under 15 or over 64 to the working population (from 15 to 64), 3) the percentage of population between 0 and 14, 4) the percentage of taxes in GDP, 5) constant per capita Gross Domestic Product, 6) military personnel as a percentage of total labor force, 7) the proportion of the central government's debt in GDP, 8) the proportion of social contribution in GDP and 9) the lack of global freedom. The instrumental variables used for corruption where: latitude and Freedom index. Her results showed that corruption significantly reduced the proportion of government spending going towards education, healthcare and social protection. The results also indicated that a higher corruption level increased the proportion of public expenditure going to public services, law enforcement, cultural activities, as well as fuel and energy. There was a significant positive relationship between corruption and defense spending. However, this relationship disappeared when freedom index was added as a control variable.

In addition to the effect corruption may have on the proportion of government expenditure going to healthcare and education, Gupta et al. (2000) found that corruption has significant negative effects on healthcare and education performance. They performed a cross-sectional regression on 128 countries using different indices based on surveys as the corruption variable. The main index used as the Political Risk Services/International Country Risk Guide. To control for endogeneity, Gupta et al. used instrumental variables for corruption identified by Treisman (2000). The instrumental variables used were share of protestants in the country, logarithm of income per capita and the exposure to democracy. Their results showed that infant mortality, percent low birthweight babies, drop-out rates in high school and especially child mortality was significantly increased by corruption. Public health and education spending was used as a control variable but was not significant in any regression apart from high school drop-out rate. The policy implication drawn by the authors from their results is that improvement of the quality of healthcare and education is possible without increasing government spending if corruption is lowered. Gupta et al. (2000) also tried a panel data approach with both fixed- and random effects. Only child mortality rates remained significant with both models, whereas the database for education was too limited for the model.

Cordis (2014) also used OLS and an instrumental variable approach when analyzing corruption's effect on public spending. However, instead of a cross-country study, the different states in the US were analyzed. The corruption variable was defined as the rate of criminal convictions of public employees for official misconduct or misuse of office similarly to the definition used by Nice (1986). Corruption was instrumented by 1) age of the state constitution as of the year 1970, 2) the number of days that an individual had to be in residence in the state as of 1970 to be eligible to vote and 3) an index of state campaign finance restrictions as of 1970. The control variables were the 1) log of real GDP per capita, 2) the percentage of state population aged twenty-five and older with a high school diploma and 3) the percentage of state population younger than 18 years. The results showed that corruption lowered the share of state spending going towards higher education and some welfare programs, while increasing spending toward unallocable budget items. However, simultaneously there was some evidence that the share of public expenditure dedicated toward healthcare was higher in US states with a higher corruption level.

3. Aims and Hypothesis

This essay aims to answer the question: Does corruption in the public sector affect the share of public spending allocated towards the selected welfare policies? More specifically, does it affect public expenditure on healthcare spending, education spending and expenditure on social services? To accomplish this, the method of fixed effects panel data regression will be used. As previously stated, the data on corruption will be collected from Transparency International's CPI.

The hypothesis explored is that corruption will be a statistically significant variable in all three of the panel data regressions using each of the dependent variables. This is because corruption has been shown in previous studies to have a negative effect on the expenditure on some welfare policies. Delavallade (2006) found evidence that corruption decreases expenditure on healthcare services. This is also supported by Mauro (1998) to a lesser extent. Addtionally, Delavallade (2006), Mauro (1998) and Cordis (2014) found evidence that education spending is negatively impacted, and Cordis achieved similar results for social protection spending. However, results differ and some previous research point to other conclusions. Even so, it seems credible that corruption would lead to programs benefiting the poor be deprioritized in favor of programs intended to create opportunities for rent seeking and benefitting a small elite.

4. Data

In this analysis, the selection of relevant welfare spending variables to be included in the analyses were based on which variables had been included in previous research on this topic, and in combination with the completeness of data available.

The variable for corruption has been collected from the Corruption Perception Index (CPI) by Transparency International, an organization working to combat and measure corruption on an international level. The CPI scores range from 0 to 100, with 0 being the most and 100 being the least corrupt (Transparency International 2021). The index was standardized in 2012 meaning that the scores only can be compared with each other from that year and onwards (Ibid). Therefore, all data collection was done from the year 2012 and onwards since corruption is the main variable of interest in this analysis.

Initially the aim was to have healthcare spending, education spending and social program spending as shares of government expenditures as dependent variables. However, there is not sufficient data available for the share of government expenditure on healthcare spending. Instead, healthcare as a share of real GDP was used as a proxy for the share of government expenditure on healthcare goods and services. Data for education spending as share of expenditure was available however and includes all education spending. The social program spending variable uses the "subsidies and other transfers" dataset collected by the World Bank. This dataset includes "subsidies, grants, and other social benefits include all unrequited, nonrepayable transfers on current account to private and public enterprises; grants to foreign governments, international organizations, and other government units; and social security, social assistance benefits, and employer social benefits in cash and in kind" (World Bank 2022a). While being the weakest of the datasets, since it does not only include spending on social programs, it is the closest to worldwide data on social programs available.

Most of the variable data was collected through the world bank's database from the years 2012 to 2019. This data in turn comprises of other databases collected by the World Bank. Sources referred to below are given by the World Bank:

<u>Dependent variables:</u>

- The healthcare variable data was collected by the World Health Organization (WHO, 2022).
- The education variable data comes from the UNESCO Institute for Statistics (2022)
- The social spending data was retrieved by the International Monetary Fund and the Government Finance Statistics Yearbook (2022).

<u>Control variables:</u>

- GDP per capita PPP adjusted in international dollars was collected by the World Bank (2022b)

- People aged between 0-14 estimated by the World Bank (2022c) based on the United Nations Population Division's age distributions.
- Age dependency ratio, the share of population in non-working age (0-14 and 65+) divided by the working age population from the United Nations Population Division (2021).
- Urbanization rate also from the United Nations Population Division (2021)

The final control variable was the Freedom Index created by Freedom House (2022). The freedom index has been used in previous research by Mauro (1995), Gupta et al. (2001), and Delavallade (2006) to control for the effect of authoritarianism as well as institutions counterbalances to corruption. The GDP per capita, previously used in most studies, and urbanization rate variable, used by Delavallade (2006) and Ali & Solarin (2019), works as a control for the level of development between nations. Additionally, PPP adjusted GDP per capita was chosen to control for differences in exchange rate and price level. Young population and age-dependency was selected specifically because they could influence education spending and has been used in previous research Delavallade (2006), Mauro (1998) and Cordis (2014). Age-dependency may also affect the healthcare variable, since it includes share of people over 65, and a large share of older population may increase healthcare spending.

5. Method

The main method used to test the hypotheses given above is a two-way fixed effects panel data regression model. The reason for using panel data regression is that it is a powerful method that allows us to capture the effect of corruption on the dependent variables over time within each country (Baltagi 2013, p.6-8). Although, this method has not frequently been applied to analyze the effect of corruption on the composition of public expenditure. There are however more recent data available, that previous studies did not have access to. Panel data assists in controlling for heterogeneity among individuals (i.e. countries) and thereby avoiding biased results. This is because the model includes individual specific effects as a term (Baltagi 2013, p.6). This term captures factors that can be considered time-invariant such as cultural, religious, and geographical differences between countries, as well as state-invariant factors that affect all nations simultaneously. This is not easily achievable when using time-series and cross-sectional data (Ibid). Additionally, panel data is more informative, with additional variability and less collinearity (Baltagi 2013, p.7). Panel data is also very effective when looking at changes in variables and identifying relationships that cross-sectional and time-series data on its own might miss (Baltagi 2013, p.8).

5.1 Panel data regression

As previously stated, this analysis uses panel data which can be described as a data set that is a collection of both cross-sectional and time-series data. This means that for several individuals (i.e. countries), data is collected for multiple variables over more than one year. As indicated above, the scores given for the corruption perception index (CPI) were only comparable from the year 2012 and onwards, therefore data collection of the other variables started year 2012. Meanwhile, there were few observations available for the year 2020. Hence, the panel data regression analysis included data for 8 years, i.e. from 2012 to 2019, including 6 explanatory variables in each regression of which 5 are control variables.

The version of linear panel data regression used in this study to analyze the effect of corruption on welfare spending is the fixed effects model. The basics formula of a two-way linear panel data regression with one explanatory variable looks like this:

$$y_{it} = \alpha + X'_{i,t}\beta + \varepsilon_{i,t} + \lambda_t$$

where the subscript *i* represents the cross-sectional dimension and *t* represents the time-series dimension (Baltagi 2021, p.358). The $X_{i,t}$ is the explanatory variable and $\varepsilon_{i,t}$ is the error term, while the β-coefficient represents the effect of the explanatory variable and α is the constant. In both the fixed effects and the random effects model that is used for panel data regression we assume individual specific effects to our error terms. With individual effects the error term is split into a random error and an individual specific effect similarly to the formula below:

$$\varepsilon_{i,t} = \mu_i + \nu_{i,t}$$

Where μ_i represents the individual specific effects and $\nu_{i,t}$ represents the part of the error term not explained by μ_i (Ibid). ν_i is most often considered homoscedastic and not autocorrelated. Since each individual in this analysis are countries, the individual specific effect to a large extent is created by the differences between nations. Countries are heterogenous due to historical, geographical and cultural reasons among others (Baltagi 2013, p.6-7). One advantage of panel data is that the individual effect is captured by μ_i instead of affecting the accuracy of the model (Baltagi 2021, p.359-361). λ_t is the time-specific effect, which reflects external events happening at a specific time (Baltagi 2013, p.39), e.g. a worldwide pandemic or economic crisis. In my case the time-specific effects can be observed by a time dummy variable for each available year.

The panel data model in this thesis is used to perform three sperate regressions. One for each of the independent variables, which are healthcare, education, and the social variable. The same explanatory variables are used in all three regressions. Namely, corruption and the control variables. In addition, the model uses differenced variables, described in section 5.4.1.

5.1.1 Limitations of Panel Data

There are nevertheless limitations of panel data regression, which are important to acknowledge. Data collection problems being one of them. Collecting a complete data set over several years with multiple variables and 178 countries is difficult. My panel is unfortunately unbalanced since it includes missing values. For some countries data for multiple years where missing and some lacked data completely for certain dependent variables. It seemed that for some smaller nations, especially small island nations, as well as in some developing nations there were more datapoints missing. This might bias the results towards the effect that can be found among larger and more developed countries. However, the data was still available for most countries. Lastly, cross-sectional dependence is a common issue among macro-panels that can affect inference.

5.2 The Fixed Effects Model

If the explanatory variable is endogenous with respect to the individual specific effect, it is known as a fixed individual specific effect. This means that if α_i is correlated with $X_{i,t}$ then fixed effects are used, otherwise a random effects model should be used (Baltagi 2021, p.363). The fixed effects estimator of β is an estimator of $\frac{dE(y|x,\alpha)}{dx}$ while the random effects estimator is an estimator of $\frac{dE(y|x)}{dx}$ (Ibid). Since it is likely that the individual effects will be impacted by the explanatory variables in the model, the fixed effects estimator was used in my model. Another reasons as to why a fixed effects model was used is that it is an appropriate model when a specific number of individuals is specifically focused on and not a random selection from a larger population. In my case I have aimed to collect data from 178 countries, i.e. all countries for which data were available. The inference in my model can only be applied to these countries. (Baltagi 2013, p.14).

The β-parameters in a fixed effects model cannot be estimated through regular OLS. They may be estimated by using a fixed effects estimator. One of the assumptions of the fixed effects model is that the sum of the $\mu_{i,t}$ error terms is zero, otherwise the estimates will be biased. The estimator first uses the within-transformator to apply de-meaning:

$$y_{i,t} - \bar{y}_i = \left(X_{i,t} - \bar{X}_i\right)\beta + \left(\alpha_{i,t} - \bar{\alpha}_i\right) + (\mu_i - \bar{\mu}_i) \;\rightarrow\; \ddot{y}_{it} = \ddot{X}_{i,t}\beta + \ddot{\alpha}_{i,t}$$

Where $\bar{y}_i$, $\bar{X}_i$ and $\bar{\mu}_i$ represents the respective means over all time periods (Baltagi 2021, p.359-360). The purpose of this is to remove the individual fixed effects because $(\mu_i - \bar{\mu}_i)$=0, since the individual effect is time constant. The fixed effects estimator $\hat{\beta}_{FE}$ is later obtained through ordinary least squares (see 5.3) using the equation above (Ibid). This estimator is the coefficient later reported in the results (see 6.) and it is an estimation of how much one unit change in the corresponding explanatory variables, changes the dependent variable. The procedure works the same if using multiple variables, when more than one $\hat{\beta}_{FE}$ is estimated.

5.3 OLS

At the core of the fixed effects panel data regression model (described in 5.2) is ordinary least squares (OLS). It is used to estimate the fixed effect estimators in the panel data regression. Generally, OLS is a way of estimating the parameters in a linear regression model by the principle of least squares (Baltagi 2021, p.175-177). This principle minimizes the sum of squares of the observed values of the dependent variable, that is: in this case the variables healthcare, education and social spending, and the predicted values of the linear function created by the explanatory variables (corruption and the control variables). The OLS estimators for the constant alpha and coefficient beta are then calculated from this formula:

$$\widehat{\alpha}_{OLS} = \bar{Y} - \widehat{\beta}_{OLS}\bar{X} \text{ and } \widehat{\beta}_{OLS} = \sum_{i=1}^{n} x_i y_i / \sum_{i=1}^{n} x_i^2$$

In a linear regression model with random sampling, the OLS estimators are the Best Linear Unbiased Estimator (BLUE) if the Gauss-Markov assumptions are upheld (Baltagi 2021, p.177-179). Those assumptions being, that the explanatory variables are exogenous, and the error terms are homoscedastic (Ibid).

5.4 Econometrical Issues

5.4.1 Non-stationarity

Stationarity is one of the requirements needed for my model to capture the effect of corruption due to the fact that inference can be affected negatively if one or more variables has a unit root and is non-stationary (Baltagi 2021, p.449). Since the panel data collected has a relatively few number of time-periods, as a consequence of the restricted time frame in the corruption index data, testing for unit roots is more difficult. This is made even harder since the panel used is unbalanced and certain observations are missing. It is however likely that some part of the collected data is stationary since the CPI does not usually display large changes between two years. In order to address the risk for non-stationary data, I have transformed my data into differences, i.e. yearly changes, through the “d. operator” in Stata. This is one used method to eliminate unit roots as well as autocorrelation (Baltagi 2021, p.136). However, it does remove one of the time periods which reduces the dataset somewhat.

5.4.2 Simultaneity

One of the biggest issues when analyzing the effect of corruption on the composition of public spending is simultaneity bias. Most likely the causality is not only in the direction from corruption to public spending, but public spending may reduce or increase opportunities for corruption to take hold in a country. Mauro (2002) describes this problem and concludes that it likely goes in both directions, meaning simultaneity most likely is present. One method of trying to remove this effect is through instrumental variables for corruption. Mauro’s use of instrumental variables for corruption has been criticized by Stephenson (2014) however.

5.4.3 Heteroscedasticity

For inference to be accurate when using a panel data regression, the standard errors must be consistent, meaning that the GM-assumptions must be upheld. One of these assumptions are homoscedasticity, meaning that the true variance of the error terms given the explanatory variables are constant over countries and over time. In this case this is not very probable stemming from size differences in economy and population between nations impacting the variance in government spending among other differences. A heteroscedasticity test (white's test) was performed on each regression, which rejected the null-hypothesis, showing that the regression indeed suffers from heteroscedasticity. The way to get around this issue and to still have consistent standard errors is to use robust standard errors which was used in my regressions analyses.

6. Results

6.1 Descriptive Statistics

From Corruption Perception Index for 2021 it becomes clear that the least corrupt countries in general are to be found in Northern Europe, North America, and richer former British colonies. All the top ranked countries would be considered rich countries by the definition of the World Bank (Hamadeh 2021). These countries all rank highly when it comes to quality of healthcare, education, and social support systems. The Nordic nations, for instance, are well known for their large welfare systems and score highly in the CPI, as shown in Table 1. On the contrary, the most corrupt nations are generally found in Africa, Central Asia as well as the Middle East. North Korea and Venezuela are stands out in their respective regions (Table 1). Many of the countries that scores the lowest are in active conflicts and most of them are poorer developing nations with authoritarian governments.

Table 1. The 10 least and the 10 most corrupt nations year 2021 according to Transparency International's (2022) Corruption Perception Index (CPI). Possible score ranges from 0 to 100, where a high value indicates low corruption.

Country	Rank	CPI score	Country	Rank	CPI score
Denmark	1	88	Turkmenistan	169	19
Finland	1	88	Equatorial Guinea	172	17
New Zealand	1	88	Libya	172	17
Norway	4	85	Afghanistan	174	16
Singapore	4	85	North Korea	174	16
Sweden	4	85	Yemen	174	16
Switzerland	7	84	Venezuela	177	14
Netherlands	8	82	Somalia	178	13
Luxembourg	9	81	Syria	178	13
Germany	10	80	South Sudan	180	11

Table 2. Summary statistics of all variables used in the regression for the period 2012-2020. All numbers are in percentage numbers, except for gdpcap. For explanation of variables, see section 4.

Variable	Summary statistics					
	Obs	**Mean**	**Std. dev.**	**Min**	**Max**	**n**
corruption (CPI)	1393	43.01	19.43	8	92	178
healthcare	1383	6.31	2.6	1.26	20.41	168
education	980	14.66	5.03	0.83	37.52	147
social	927	40.9	19	2.01	84.61	124
gdpcap	1372	19897	20812	670	141635	173
youngpop	1400	28.05	10.8	11.05	50.26	175
agedep	1400	59.31	17.92	16.14	111.94	175
urbanpop	1408	58.64	22.47	11.19	100	176
freedom	1416	57.98	29.25	-1	100	177

As indicated in Table 2, the corruption variable CPI has a mean of 43 out of all observations for the studied period 2012-2020. The very wide range, being 8 to 92, illustrates the magnitude of the issue of corruption. The CPI variable is also illustrated in the boxplot in Figure 1, showing that the median has a score of 38, which is slightly lower than the mean of 43. It is also clear that the social variable has a very large span, ranging from 2.0 to 84.6 (Table 2). The two variables with the fewest number of observations by a noticeable margin is the education and social variables. Both these variables are given as shares of government expenditure. It seems to be more difficult to retrieve data related to government expenditure than data of shares of GDP, as indicated by the healthcare variable that has more observations.

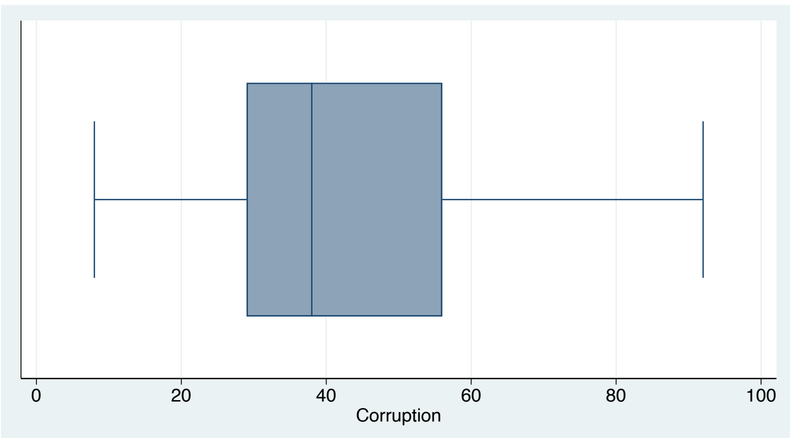

Figure 1. Boxplot over the distribution of the corruption variable for 1393 observations over 178 countries. The line in the middle of the box visualizes the median of all observations of the Corruption Perception Index. The

edges of the box represent the observations found at the 25th and the 75th percentile. The left whisker represents the minimum value, and the right whisker illustrates the maximum value.

In Figure 2, the distribution of the three dependent variables included in the panel data regressions are shown. Healthcare spending as a share of GDP spans between 1 and 13 percent, with some outliers having a larger share. Education spending out of government expenditure have an even larger span, ranging between 1.5 and 27 percent with multiple outliers. The median of the share of education spending is approximately 14 percent. As indicated above, the social variable has a very large span. The reason for this could be due to actual differences in social spending, but the fact that different nations report their statistics differently might contribute. For example, different governments could classify different programs as social programs.

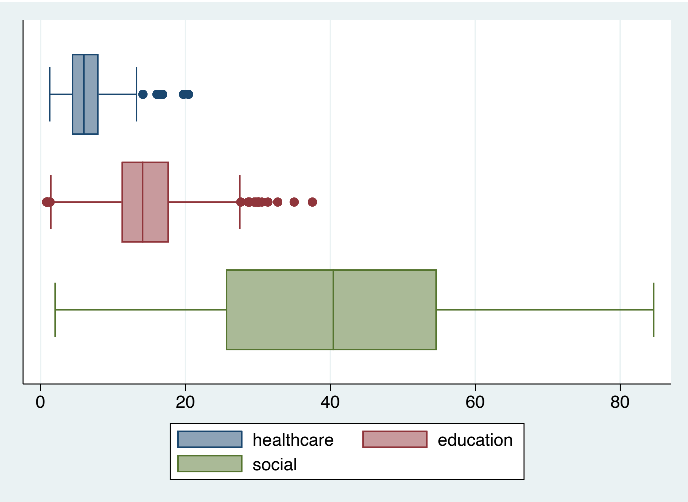

Figure 2. Distribution of the three dependent variables being analyzed. Upper boxplot illustrates the distribution of countries healthcare spending as shares of GDP. The middle boxplot represents the shares of government expenditure on education and likewise the shares of government expenditure spending of social service, transfers and subsidies is shown in the lower boxplot. Outlier observations outside of 1,5 times the distance between the 25th and 75th percentile is represented by dots.

6.2 Fixed effect results

In Table 4 the regressions results are displayed for the three dependent variables. The year dummy variables from 2014 to 2019 represents the λ_t time-specific effects of the model for each year. Firstly, when observing healthcare as a dependent variable, corruption is significant with a p-value of 0.027. The coefficient is relatively low at 0.0208, nonetheless the result supports the hypothesis that corruption negatively impacts the public expenditure towards healthcare. Interestingly, the only control variable that proved significant was the logarithm of GDP per capita with a negative coefficient. An increased GDP per capita seems to lead to less spending on healthcare per dollar unit. Healthcare may in that sense be likened to an inferior good. It could also be that some allocation of the economy must be directed towards healthcare since it provides a basic need leading poorer nations to spend a larger share of GDP on these services. The only statistically significant time-specific effect occurred during 2017 with a negative coefficient of 0.178.

Regarding the regression analyzing government education spending the corruption variable proved to be almost significant at the 5 per cent level at a p-value of 0.059 and a positive coefficient of 0.0878. This time the ln GDP per capita variable is not significant but the share of children 0-14 years of age and the age dependency ratio. It is noteworthy that the youngpop variable coefficient is -2.714 while the age dependency ratio is positive with a coefficient of 1.155. This suggest that more developed nations have fewer children on average and hence have more resources to spend on education. The significant time-effect occurred the year 2019 with a negative coefficient of -0.88.

Table 3. This table shows the results of the fixed effects regressions using all available data including the control variables. The coefficients are the β_{FE} estimators of each explanatory variable. The number of countries used in each regression are shown in parenthesis next to the number of observations.

Variable	**Healthcare**			**Education**			**Social**		
	Coefficient	**Robust SE**	**P-value**	**Coefficient**	**Robust SE**	**P-value**	**Coefficient**	**Robust SE**	**P-value**
Corruption	0.0208	0.00929	**0.027**	0.0878	0.0462	0.059	-0.0293	0.175	0.867
Freedom	-0.00716	0.00743	0.336	0.0186	0.430	0.666	-0.0542	0.0721	0.454
Urban pop.	-0.191	0.383	0.619	0.541	1.22	0.659	-10.3	6.70	0.126
Age Dep.	-0.118	0.163	0.470	1.16	0.533	**0.032**	-0.685	1.78	0.701
Young pop.	0.365	0.384	0.344	-2.71	1.219	**0.028**	4.01	4.84	0.409
Ln GDP/cap	-1.12	0.448	**0.013**	-4.63	3.033	0.129	9.04	4.04	**0.027**
2014	-0.0531	0.792	0.503	-0.343	0.406	0.399	0.169	0.563	0.765
2015	-0.0558	0.0602	0.355	-0.207	0.313	0.510	0.272	0.603	0.653
2016	-0.0889	0.0533	0.098	-0.272	0.290	0.349	-0.109	0.600	0.856
2017	-0.180	0.0590	**0.003**	-0.0487	0.340	0.887	-0.435	0.525	0.409
2018	-0.0920	0.0603	0.129	-0.334	0.309	0.282	-0.163	0.655	0.804
2019	-0.0766	0.0595	0.200	-0.881	0.331	**0.009**	1.22	2.65	0.645
Within R^2	0.0275			0.0293			0.0092		
Obs. (n)	1124 (168)			757 (147)			738 (124)		

Furthermore, using the social variable as a dependent variable in the regression did not give a significant p-value for the corruption variable. Neither did any of the control variables except ln GDP per capita with a very large coefficient of 9.045. No significant time-specific effects occurred either. The F-statistic for this regression is 0.662 which could indicate that the model is miss-specified, explaining the low R_2-value. Regarding the previous 2 regressions the F-statistic was 0.0039 when healthcare is the dependent variable and 0.0117 when it is education.

In order to make a refined analysis, fixed effects regressions were made on different segments of countries. In Table 4, regressions were made for different countries classified according to their levels of GDP per capita, enabling the analysis on how the p-value and coefficient changes between different sized economies. In the lower income subset, the corruption received a very low p-value for healthcare, although very few observations were available. Only four countries in the data set belonged to this class and the result is therefore considered uncertain. Lower middle-income had a p-value slightly below 10% for the healthcare variable, whereas the richer subsets of countries both had a p-value way above the significance level ($p>0.2$). It is when combining low- and middle-income levels that corruption becomes a significant variable for healthcare with a p-value of 0.023. Hence, by splitting the countries into two approximately equal groups (96 and 89 countries respectively), it comes apparent that it is in the group representing the poorest countries where the relationship between corruption and healthcare spending is strongest.

Just as in the case for healthcare, the corruption variable has a low p-value ($p=0.000$) in the low-income subset of education. Also, here there are very few observations, and the result must be considered as uncertain. Corruption is almost statistically significant at the 5% level for the upper-middle-, high- and the combined upper-middle- and high-income brackets together. Surprisingly, corruption in the lower-middle income countries received a p-value of 1. In contrast to healthcare, the relationship between corruption and education is strongest for the upper-middle income and the richest half of the countries (not counting the low-income subset), being almost significant ($p=0.054$ and $p=0.059$ respectively).

Table 4. Fixed effects regressions using the same variables as in table 4 but only reporting for the corruption variable and looking at different income levels. These levels were decided using the World Bank's definitions of country income levels (World bank, 2022d). Low income= GDP per capita<1046, Lower-middle income= 4096>GDP per capita>1056, Upper-middle income= 12696>GDP per capita>4095, High income= GDP per capita>12695.

Income levels	**Healthcare**				**Education**				**Social**			
	Coefficient	**Robust SE**	**P-value**	**Obs. (n)**	**Coefficient**	**Robust SE**	**P-value**	**Obs. (n)**	**Coefficient**	**Robust SE**	**P-value**	**Obs. (n)**
Low income	0.564	0.0281	**0.000**	19(4)	1.75	4.45E-07	**0.000**	14(4)	-	-	-	-
Lower-middle income	0.0471	0.0280	0.099	247(44)	-3.12E-05	0.109	1.000	184(41)	-0.133	0.562	0.814	148(30)
Upper-middle income	0.0148	0.0123	0.233	290(54)	0.186	0.0938	0.054	184((42)	3.94E-04	0.179	0.998	173(40)
High income	0.00573	0.00812	0.482	568(96)	0.107	0.0551	0.061	375(81)	-0.104	0.233	0.657	428(73)
Low & middle income	0.0382	0.0170	**0.027**	556(89)	0.0956	0.0677	0.162	382(78)	-0.130	0.274	0.636	327(63)
Upper middle & high income	0.00896	0.00700	0.203	858(133)	0.104	0.0546	0.059	559(111)	-0.0491	0.170	0.774	601(104)

Table 5. The results of fixed effects regressions using all control variables for subsets of different levels of corruption.

CPI score	**Healthcare**				**Education**				**Social**			
	Coefficient	**Robust SE**	**P-value**	**Obs. (n)**	**Coefficient**	**Robust SE**	**P-value**	**Obs. (n)**	**Coefficient**	**Robust SE**	**P-value**	**Obs. (n)**
CPI<26	0.00325	0.0191	0.866	159(38)	0.0198	0.139	0.888	96(29)	-1.0523	0.526	0.058	69(23)
51>CPI>25	0.0279	0.0135	**0.041**	619(105)	0.0946	0.0860	0.277	227(49)	0.0217	0.206	0.917	391(75)
76>CPI>50	0.0110	0.0129	0.396	240(46)	0.130	0.101	0.204	161(38)	0.0995	0.412	0.811	191(39)
CPI>75	0.00213	0.0106	0.844	106(18)	-0.0455	0.0646	0.492	73(16)	-0.161	0.407	0.698	104(17)
Corruption<51	0.0294	0.0127	**0.023**	778(122)	0.0744	0.0584	0.205	523(105)	-0.191	0.204	0.353	460(85)
Corruption>50	0.0110	0.0103	0.291	346(60)	0.0923	0.0838	0.276	234(51)	0.0753	0.322	0.816	295(53)

Regarding the social variable in low-income countries there were too few observations to do a panel regression. Additionally, for all levels of income corruption was statistically insignificant reflecting the results using the entire dataset (table 4).

In a final data analysis, the countries were grouped into different quartiles of corruption (Table 5). For healthcare, corruption is significant for the group of countries having a CPI between 51 and 25. Furthermore, the significance increased when all countries with a CPI below 51 (n=122) were grouped together (p=0.023). Notably, no subset of corruption level for education resulted in a p-value below 0.2. Apparently, for education the correlation with corruption is stronger when all the countries are included in the analysis. Meanwhile, the subset of the social variable with a CPI under 25 had an almost statistically significant p-value of 0.058, with 69 observations over 23 countries. If there is a correlation between social spending and corruption it might be here. The F-statistic for this regression was near 0 in contrast to when running the regression on the whole dataset.

7. Discussion

One major result of the panel data analyses was that the healthcare variable had a significant (p=0.027) negative relationship with corruption (Table 3). A lower score on the CPI – meaning a higher perceived corruption level– coincides with a lower share of GDP directed towards healthcare related goods and services, when using all available observations. However, when grouping countries based on GDP per capita, this effect is not noticeable among richer nations, but solely for the low- and middle-income subsets (Table 4). Similarly, corruption only remained a significant variable in the subsets where the CPI scored under 51 and between 25 and 51 (Table 5). Even if the effect of corruption on the share of healthcare expenditure only occurs for countries with at high levels of corruption, this is a large segment of the world's nations; approximately two thirds of all countries received a scoring below 50. The β_{FE}-coefficient is nonetheless relatively low, 0.021 using all observations (Table 3), or 0.029 for the group with a CPI-score below 51 (Table 5). Since differenced variables have been used in this analysis the coefficient has two meanings: Firstly, one point increase in the CPI increases healthcare spending by 0.021% percentage points. Secondly, one point change in growth of the CPI increases the change of growth in healthcare spending by 0.021% percentage points.

The result that the share of healthcare spending decrease as corruption increase is a further confirmation of the findings of Delavallade (2006). She included 51 developing and 13 developed countries with data for the period 1996 to 2001 in her study. Not only the period and selection of countries differed from my study, but also the corruption indicator that was taken from the World Bank (see section 2.1). However, this result is not in agreement with the results of Cordis (2014) who found that healthcare spending was somewhat higher in more corrupt states in the USA. However, Cordis' analysis of the US' states may not only apply to other countries since the US does have a relatively unique healthcare system dominated by private insurers and is an outlier among world nations when it comes to healthcare spending per capita (much higher).

There was a weak, nearly significant relationship (p=0.059) using education as the dependent variable (Table 4). This value may not be under the significance level of 5%, nonetheless it is very close to be significant. While being significant or nearly significant in three subsets of GDP per capita, it remained insignificant when looking at different subsets of CPI. Overall, the results align with results presented in multiple research articles (Mauro, 1998; Delavallade, 2006; Cordis, 2014). Mauro (1998) analyzed the effect of corruption on government spending on defense, education, transfers, social security and welfare, and total government consumption expenditure for about 100 countries with data for the period 1970–1985. He found that government spending on education as a ratio to GDP was negatively and significantly correlated with corruption. In my analysis, the coefficient remains relatively low, although higher than for healthcare, with a value of 0.088 when using all observations. The coefficient increases when looking at the significant subsets of GDP per capita, ranging from 0.105 to 0.186, not including the low-income subset due to few observations. If true, an improvement of 10 points on the CPI would result in a 1.8 percentage point increase in education spending in an upper-middle income country as share of government expenditure. In reality, this could be a large amount of money distributed towards education.

No relationship was found between CPI and the social variable. Not only is the corruption variable not significant, but all explanatory variables also used are insignificant and so is the F-statistic. This could mean that the hypothesis is incorrect, and that corruption does not lead to a decrease in social expenditure or that the data is too unreliable to use in a regression. The social variable data include transfers and subsidies, which could impact the regression.

Additionally, this variable is not as standardized as the other two, risking different interpretations among countries of what the data should be included. This may also explain why the social variable is so widely distributed, as illustrated in Figure 2. The variable also include data that is not social program related, such as intergovernmental transfers and subsidies to private companies. To conclude; it is hard to draw solid conclusions from the results concerning the social variable due to the high degree of uncertainty in this variable.

The results presented in Tables 4-6 suggest that corruption indeed may influence certain types of welfare spending negatively in the form of healthcare spending and education spending shares. Following the reasoning of Mauro (1998), based on the research of Kreuger (1974), corruption will occur in government sectors that easily allows for rent-seeking and will increase funds directed towards these government sectors. Healthcare and education sectors of government might not be accessible to rent-seeking behavior to the same extent as other sectors. Furthermore, if this is true corruption pose a threat to healthcare and education services provided by the government, since these areas will lose funding redirected to other areas. The implication of this is that efforts aiming to lower corruption could also lead to a larger public expenditure share allocated to healthcare and education. Apart from improving the standard of living, these efforts could also increase economic growth in the long run since a functioning education and healthcare system is associated with higher economic growth.

There were some potential sources of inaccuracy in the data. A main one being that the panel data is unbalanced. This may not be a problem if only random observations are missing. However, if this is not the case there is a risk that there is a non-random effect added to the error term in the regressions. In the database used it seems like smaller and poorer nations are more likely to lack data in general, even though this was not always the case. This means that there may exist an additional component to the error-term that might influence the result. Nevertheless, a large amount of data was retrieved in this study and even the social variables had data collected for 124 countries, i.e. most of the world's nations. Additionally, a potential error source could be the CPI itself. As is described in section 2.1, there are some limitations when measuring the perception of corruption. The main critique of the CPI is that corruption is hard to measure and that lags between corruption and the perception of corruption may occur. Even so, the use of corruption indices has become more widely used simultaneously as more data has been collected. Villarino (2021) concluded that both the CPI and the World Bank's Control of Corruption Index was somewhat reliable and passed the bar of minimum validity. This is supported by the agreement between the results obtained by Delavallade (2006) with the World Bank indicator of corruption and the Transparency International's CPI index used in the present study, regarding the link between corruption and healthcare spending.

An econometrical problem present in my analysis is the risk of simultaneity bias. Mauro (2002, p.349-359) describes how the effect corruption on the composition of government expenditure most likely goes in both directions and that assumptions about causality should be made cautiously. He and others have tried to compensate for this by using a 2-stage or 3-stage least square approach instrumenting for corruption with several instrumental variables. Nonetheless, this use of instrumental variables has received criticism (Stephenson, 2014) for not using variables fulfilling the requirements of a good instrumental variable. Due to this difficulty to account for simultaneity, the focus of the present study was on the panel aspect of the data. A suggestion for future research would be to find a way to account for the possibility of simultaneity, either by attempting to find sufficient instrumental variables and using an instrumental variable approach or other methods such as using lagged data.

One aspect to consider when analyzing corruption's effect on policies like healthcare, education, and social spending, is that the budget dedicated towards these programs, most likely also is affected by the political philosophy of the ruling party (or parties). Such a variable is not included in my model since it is a complicated one to construct. Even though, the freedom index variable might reflect some aspects of the political philosophy of a nation. This means that the effect political philosophy has on the budget share of welfare policies will be added to the error-term in the model. Part of this effect might be captured by the individual specific effects of a country, which is one advantage of panel data. Even so, this likely affects the efficiency of the model and may contribute to some of the regressions having a low R^2 value. A suggestion for similar analysis in the future would be to use some control variables reflecting political philosophy, such as left- or right-leaning ruling party.

This thesis is one of few articles using a fixed effects regression to analyze to the effect of corruption on the composition of welfare expenditure. It also one of few to use the CPI as a basis for the corruption variable in this area. As new data is collected in terms of the CPI and other corruption indices, the possibilities for panel data analysis of corruption will improve. Similarly, when more data is collected on government expenditure it will facilitate using healthcare spending as a share of public expenditure rather than GDP. Other suggestions for future research would be to also analyze the effect of corruption on government spending in areas that might be more prone to rent-seeking e.g., infrastructure or military spending or more types of welfare spending. As more data become available a fully balanced panel data analysis with all nations may also be possible. It would be interesting to see if the results such an analysis will differ from an unbalanced analysis.

8. Conclusion

In conclusion, the fixed effects regressions performed in this study indicates that corruption negatively affects healthcare spending as a share of GDP. This adds support to the hypothesis that corruption lowers the share of healthcare in government expenditure. This effect seemed most significant in low- and middle-income countries as well as countries with a CPI score under 51. The results also somewhat support that corruption negatively influences share of government spending on education with a p-value close to significant. These findings are supported by some of the previous literature on the subject. However, corruption could not be shown to affect the share of expenditure on social spending, even though some previous research support such a relationship. This may be caused by lacking standardization, as well as broad definition of the dataset used to analyze social spending. These results may be affected by simultaneity bias making it hard to show direction causality between corruption and healthcare expenditure as well as corruption and education expenditure. Lastly, there are broad possibilities for future research on this topic with corruption indices improving and the collection of new data.

9. References

Baltagi, Badi, 2013. *Econometric Analysis of Panel Data*. 5th Ed. Cornwall: John Wiley & Sons Ltd.

Baltagi, Badi, 2021. *Econometrics*. 6th Ed. Syracuse, NY, USA: Springer Nature Switzerland AG.

Bello y Villarino, José-Miguel, 2021. ”Measuring Corruption: A Critical Analysis of the Existing Datasets and Their Suitability for Diachronic Transnational Research.” *Social Indicators Research,* no. 157, pp.709–747 (2021).

Brittanica, 2022. *Rent Seeking.* [Electronic] https://www.britannica.com/topic/rent-seeking. Retrieved: 27-06-2022.

Cordis, Adriana S, 2014. ”Corruption and the Composition of Public Spending*”, Public Finance Review*. Vol. 42, no. 6, pp. 745-773.

Delavallade, Clara, 2006. ”Corruption and distribution of public spending in developing countries”, *Journal of Economics and Finance*, Vol. 30, pp. 222-239.

De Maria, Bill, 2008. "Neo-colonialism through measurement: A critique of the corruption perception index”. *Critical Perspectives on International Business*, vol. 4 no. 2-3 pp. 184–202.

European Commission, 2020, ”Corruption”, Eurobarometer. (First published June 2020) [Electronic] https://europa.eu/eurobarometer/surveys/detail/2247. Retrieved: 27-05-2022

European Commission, 2022, ”2022 Rule of law report” (The rule of law situation in the European Union), Published July 13 2022. [Electronic] https://ec.europa.eu/info/policies/justice-and-fundamental-rights/upholding-rule-law/rule-law/rule-law-mechanism/2022-rule-law-report_en. Retrieved: 2022-07-25

Goel, Rajeev K, Nelson, Michael A. 2010. ”Causes of corruption: History, geography and government”. *Journal of policy modeling*, vol 32, issue 4, July–August 2010, Pages 433-447.

Gupta, Sanjeev – Davoodi, Hamid – Tiongson, Erwin, 2000, ”IMF working paper/ Fiscal Affairs department, Corruption and the provision of health care and education services”, *International Monetary Fund*.

Hamadeh, Nada – Van Rompaey, Cathrine – Metreau, Eric, 2021, ”New World Bank country classifications by income level 2021-2022”, *World Bank Blogs* (July 1 2021) [Electronic] https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2021-2022. Retrieved: 20-05-2022

Hawken, Angela, Munck, Gerardo L. 2014. ”Does the evaluator make a difference? Measurement Validity in corruption research”. *Political Concepts Working Paper Series # 48* International Political Science Association [IPSA], Committee on Concepts and Methods, 2011. Retrieved: 21-05-2022

International Monetary Fund, 2022. ”IMF data access to macroeconomic & financial data”. [Electronic] https://data.imf.org/?sk=a0867067-d23c-4ebc-ad23-d3b015045405 (Update July 7 2022). Retrieved: 04-04-2022

Kaufmann, Daniel, Kraay, Aart, Mastruzzi, Massimo, 2007. ”Measuring Corruption: Myths and Realities". Africa Region Findings & Good Practice Infobriefs; no. 273.

Lambsdorff, Johann Graf, 2007. ”Measuring corruption – the validity and precision of subjective indicators (CPI)”. *In Measuring corruption*, (pp. 81–100). Taylor and Francis.

Mauro, Paolo, 1995. ”Corruption and growth", *The Quarterly Journal of Economics*, Vol. 110, issue 3, pp 681–712.

Mauro, Paolo, 1998, ”Corruption and the composition of government expenditure”, *Journal of Public Economics*, Vol. 69, issue 2, pp 263-279.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, 1993. "Why Is Rent-Seeking so Costly to Growth?" *American Economic Review Papers and Proceedings*, Vol. 83, No. 2, pp.409-14.

Nice, David C, 1986, ”The policy consequences of political corruption”, *Political Behaviour*, Vol. 8, No. 3, pp. 287-295.

Stephenson, Matthew, 2014. *Mauro (1995) does not show that corruption slows growth*. [Electronic] https://globalanticorruptionblog.com/2014/05/15/mauro-1995-does-not-show-that-corruption-is-bad-for-growth/. Retrieved: 22-04-2022

Stephenson, Matthew, 2015. *The 2014 CPI Data Demonstrates Why, Even Post-2012, CPI Scores Cannot Be Compared Over Time*. [Electronic] https://globalanticorruptionblog.com/2015/02/17/the-2014-cpi-data-demonstrates-why-even-post-2012-cpi-scores-cannot-be-compared-over-time/. Rerieved: 06-07-2022

Transparency International, 2021. *The ABCS of the CPI: How the corruption perceptions index is calculated*. [Electronic] https://www.transparency.org/en/news/how-cpi-scores-are-calculated (December 2021) Retrieved: 10-04-2022

Transparency International, 2022a. Corruption Perceptions Index *(CPI) 2021 The results at a glance*. [Electronic] https://www.transparency.org/en/cpi/2021. Retrieved: 01-04-2022

Transparency International, 2022b. *2021 CORRUPTION PERCEPTIONS INDEX REVEALS A DECADE OF STAGNATING CORRUPTION LEVELS AMID HUMAN RIGHTS ABUSES & DEMOCRATIC DECLINE.* Published 22 January 2022. [Electronic] https://www.transparency.org/en/press/2021-corruption-perceptions-index-press-release. Retrieved: 04-04-2022

Transparency International, 2022c. *What is Corruption?*. [Electronic] https://www.transparency.org/en/what-is-corruption. Retrieved: 20-07-2022

Transparency International, 2022d. *Corruption Perceptions Index (CPI) 2021– Full Source Description*. [Electronic] https://images.transparencycdn.org/images/CPI2021_SourceDescriptionEN.pdf. (January 2022) Retrieved: 30-03-2022

United Nations Population Division, 2021. *UN Population Division Data Portal.* [Electronic] https://population.un.org/dataportal/home. Retrieved: 04-07-2022

World Bank, 2022a. *Subsidies and other transfers (% of expense).* [Electronic] https://data.worldbank.org/indicator/GC.XPN.TRFT.ZS Retrieved: 04-02-2022

World Bank, 2022b. *World development indicators database. International comparison program*. [Electronic] https://databank.worldbank.org/reports.aspx?source=2&type=metadata&series=NY.GDP.PCAP.PP.CD . Retrieved: 04-20-2022

World Bank, 2022c. *Population ages 0-14 (% of total population)*. [Electronic] https://data.worldbank.org/indicator/SP.POP.0014.TO.ZS Retrieved: 04-07-2022

World Bank, 2022d. *World Bank Country and Lending Groups.* [Electronic] https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups Retrieved: 04-29-2022

Zaman, Asad, 2009. ”Corruption: Measuring the Unmeasurable”. *Humanomics*, 25(2), pp 117–126.