AI-powered insights on global corruption: a multi-view analysis

Zsolt János Viharos
HUN-REN Institute for Computer Science and Control,
John Von Neumann University, Hungary
Email: viharos.zsolt@sztaki.hu; viharos.zsolt@nje.hu

Anh Tuan Hoang
HUN-REN Institute for Computer Science and Control,
Hungary
Email: hoang.laszlo@sztaki.hu

Botond Géza Kálmán
Faculty of Economics and Business,
John Von Neumann University, Hungary
Email: kalman.botond.geza@nje.hu

Endre Béla Huff
Faculty of Economics and Business,
John Von Neumann University, Hungary
Email: huff.endre@nje.hu

Zoltán Zéman
Doctoral School of Business and Organizational Sciences,
John Von Neumann University, Hungary
Email: zeman.zoltan@nje.hu

Corruption is a major global problem. It undermines economic growth, government efficiency and public trust in institutions. Therefore, predicting the evolution of the corruption situation is a very important goal. However, choosing which indicators to use is difficult. The perception of corruption either measures a subjective corruption perception index or is based on objective statistical numbers. However, these latter statistics do not show undetected cases of corruption and therefore underestimate the frequency of the phenomenon. This study uses a machine learning method to select economic and socio-statistical indicators that can be used to estimate corruption. Based on the method, the trend of the relationship between the indicators included in the research and the perception of corruption or the lack thereof enriches the knowledge on the subject with a lot of information. The results also bring science closer to developing a reliable forecasting method.

Keywords:
corruption,
artificial intelligence (AI),
country positioning,
feature selection,
background explanation

Introduction

A review of the literature proves that it is worth looking for a relationship between macroeconomic indicators and the perception of corruption. In the first and second parts of this study, the authors undertook this as well. This article presents the results of one of the paths of a two-pronged research (the impact of hard socio-economic data on the development of corruption perception). The authors describe the effects of soft data in another study, which has not yet been published. The novelty of the study is primarily the method used. The indices included in the study were selected with the help of *artificial intelligence* (AI) using an adaptive feature selection algorithm based on machine learning. The authors then analysed the data selected in this way using the *feature analysis* through graphical representation procedure and drew their conclusions. In some cases, it was possible to demonstrate a trend-like relationship between the examined economic indicators and the perception of corruption. The analysis of these is described by the authors in another study. The current study describes the results related to additional indices that do not show a trend but were selected with the applied methodology. Despite the fact that no real trend was observed for these variables, the authors marked a quasi-trend in the figures illustrating the results to facilitate the analysis. The authors prepared a short theoretical and methodological introduction, the detailed literature background and the applied methodology will be included in an upcoming article.

Literature review

Corruption and economics

Transparency International (TI) (2024) publishes the *corruption perceptions index* (CPI) of 180 countries every year. This assesses the level of corruption perceptible in the public sector (Abramo 2008, Donchev–Ujhelyi 2014). TI has already received many criticisms (Budsaratragoon–Jitmaneeroj 2020, Németh et al. 2019, 2021), and in many cases its results are not used properly. Such misuse is, for example, when journalists misinterpret the results or deliberately confuse corruption (which TI does not measure) and the perception of corruption (Előd 2020, Portfolio.hu 2021). The latter is measured by the CPI, but perception is not the same as actual corruption (Szemán 2022). Such inappropriate use of the index causes economic damage and tarnishes the reputation of the given country. Overall, the index is still widely used and an internationally accepted benchmark. That is why the authors decided to use it in their research.

The phenomenon of corruption is almost as old as history. Its strengthening can be attributed to the growth of the economic influence of the merchant bourgeoisie in the new era (Engels 2016). Engels explains the background of the statement in detail in his referenced book. However, due to the length of the article, we did not want to go into too much detail about the book. The essence of the statement is that in the pre-modern period, power and money were concentrated in the same persons (rulers, leaders). In the modern era, the situation changed. The bourgeoisie had money without political power and the ruling class had power but less and less money. This system of relations almost obviously leads to corruption in the form of power bought for money.

In this study, the authors only point out that corruption always involves obtaining personal benefits, as highlighted by the TI definition (Nagy 2018). And the advantage can always be expressed in money, even if, for example, information is obtained by the corrupt party (Kyung–Nam 2023, Wang–Steinberg 2010). The connection between corruption and the flow of money is much more obvious in connection with money laundering (Barone et al. 2022, Chaikin–Sharman 2009) and the diversion of public procurement resources (Sergi–South 2016, Walters 2013). This reduces the efficiency of the use of public funds, thereby worsening the performance of the economy (Tanzi–Davoodi 1997). Deteriorating performance then reduces the desire to invest, one of the effects of which will be an increase in unemployment due to the lack of investment. The other consequence is the downgrading of the country by the credit rating organisations due to the increased investment risk (Budik–Ezr 2015).

Many studies have dealt with the relationship between corruption and economic performance. The relationship between GDP and CPI was a frequent research topic. It has been proven that a high level of corruption reduces the incentive to invest, increases company costs and undermines the efficiency of public institutions (Atsir–Sunaryati 2018). Due to its global nature, corruption also negatively affects foreign trade indicators (Atsir–Sunaryati 2018). It plays a role in many sectors of the economy (Campos–Pradhan 2007) and even has a negative impact on competitiveness (Budik–Ezr 2015) and sustainability goals (Fhima et al. 2023).

Feature selection

In the finance sector, feature selection is essential for optimising predictive models and decision-making processes. Various financial domains such as risk management (Lahmiri 2016), stock price prediction (Peng et al. 2021), fraud detection (Ravisankar et al. 2011) and distress prediction (Liang et al. 2015) use feature selection algorithms to pinpoint the most relevant variables. These techniques aim to reduce dimensionality, enhance model interpretability and improve prediction accuracy by selecting a subset of features from the original dataset. Research in finance has shown that feature selection effectively enhances model performance, reduces overfitting and identifies key drivers of financial outcomes.

Additionally, numerous studies thoroughly explore the relationship between corruption perception and socio-economic indicators across various countries. Using advanced data analysis techniques like cluster analysis and machine learning, the CPI is examined by selecting key features and analysing their impact (Domashova–Politova 2021). By employing feature selection methodologies such as the *add* and *add–del* methods, the study strategically identifies the most influential indicators from economic, social and business management domains. Analysing data from 2012 to 2019, the research highlights the complex relationship between corruption perception and various factors, offering valuable insights for policymakers and stakeholders aiming to combat corruption and improve governance globally.

Another study investigates corruption analysis across countries using self-organising maps (SOMs) and support vector machines (SVMs) (Huysmans et al. 2006). The research aims to explore the connections between macroeconomic variables and perceived corruption levels. Initially, SOMs are used to visualise the interconnections among variables, followed by SVMs for forecasting corruption levels. For feature selection, methods like the backwards procedure with leave-one-out cross-validation error were used to identify significant variables. The study integrates data from multiple sources including the CIA Factbook and TI, and the results indicate potential deviations between predicted and actual corruption levels, providing insights for further research.

Bayesian models have also been developed to assess corruption risk in federal management units (Carvalho–Carvalho 2016). Techniques like correlation analysis and adaptive lasso were used for feature selection, with naïve Bayes chosen as the final model, achieving an AUC of about 0.82. Significant rules were extracted to identify corruption-prone units. A web application enables CGU managers to prioritise anti-corruption efforts effectively. This study demonstrates the utility of data mining in combating corruption and maximising audit efficiency.

Research on corporate anti-corruption disclosure using feature selection as a machine learning approach, compared to traditional statistical methods, was also conducted (Utomo et al. 2021). By analysing data from the Indonesian Stock Exchange and the United Nations Global Compact, the study identifies variables influencing anti-corruption policies. Feature selection methods such as *decision tree*, *support vector machine* and *random forest* were compared with statistical analysis, showing superior performance in identifying crucial variables for anti-corruption disclosure. The study highlights the efficacy of machine learning in enhancing corporate governance decision-making related to anti-corruption policies.

Methods

Dataset

The present research is based on two publicly available databases. One is the CPI published every year by TI. This ranks 180 countries based on perceived corruption in the public sector (Transparency International 2024). The ranking is based on the corruption detection score calculated using the methodology developed by TI.

The methodology developed by Lambsdorff (2008) was changed in 2012 to better adapt the results to modern expectations (Saisana–Saltelli 2012). The extent of the change is so significant that TI specifically draws attention to the fact that the scores before and after 2012 cannot be compared with each other. Therefore, the present study also uses data after 2012.

The CPI is prepared by processing questions from a total of thirteen sources investigating corruption. However, most of these sources contain not only corruption but also economic and social data. A particularly rich source of data is the *sustainable governance indicators* (SGI) database of the Bertelsmann Stiftung (2023a). This data series also has its own calculation methodology (Bertelsmann Stiftung 2022). The authors of this study chose this database because it also contains so-called 'hard data' based on statistics. Such data are generally considered more reliable than subjective ('soft data') opinions obtained through questionnaires. However, the usability of the data also depends on the topic of the research. Most approaches to corruption, including the CPI, are based on subjective opinions. Although there are also objective approaches – for example, the number of court judgments can also be examined – these generally cannot provide an estimate of the level of hidden corruption. Nowadays, there is no generally accepted and unquestionable method for examining this hidden area (Németh et al. 2019).

Beliaeva et al. (2020) concluded that subjective measurements can be more reliable, especially during crises. The big disadvantage of objective performance measures is the absolute nature of these indicators, compared with subjective performance measures (Wall et al. 2004). So, subjective performance measures solve most of the problems and limitations, especially when dealing with the confounding effect of Covid-19. They also help overcome the problem that companies may not be willing to share their own information. To verify this claim, the authors of this study conducted further research using subjective opinion data from the SGI.

The Bertelsmann Stiftung SGI have been measuring countries' sustainability and governance performance since 2009. They examine three areas: sustainability, level of democracy and economic performance (Schiller et al. 2023). They evaluate sustainability on an economic and political level, examining economic and financial indicators as well as the social system. Democracy is characterised by the rule of law, freedom of the press, political participation and the quality of democratic institutions. Government performance is described by the ability to make policy decisions and implement policy reforms in practice (Bertelsmann Stiftung 2023b). The number of countries examined is 41. Each of the countries examined is a member of the OECD, the European Union or both (Bertelsmann Stiftung 2023a). This is also a weakness of the database, since the performance of other countries cannot be determined from the data. However, there is also a relationship between corruption and economic indicators in developing countries (Uddin–Rahman 2023).

The country reports are prepared by independent experts for the Bertelsmann Stiftung. The reports also include expert assessments and statistical data from international organisations (OECD, EU, IMF) and local sources (Bertelsmann Stiftung 2022). Although the scope of the examined countries is limited, the analysis of these countries is carried out in detail according to many aspects. The performance of the countries can be compared with each other. The regular updates of the reports enable the monitoring of long-term trends and changes in each country.

Adaptive, hybrid feature selection (AHFS)

The *adaptive, hybrid feature selection* (AHFS) algorithm offers a flexible approach to feature selection in machine learning by merging existing supervised techniques and emphasising adaptability (Viharos et al. 2021). Its robustness and effectiveness make it suitable for real-world applications (Nagy et al. 2023). The algorithm integrates various correlation and information theoretic measures, such as MMIFS, mRMR and LCFS, to evaluate the relationship between features and their information content. This hybrid nature allows the AHFS algorithm to adapt to different scenarios and incorporate future scientific advancements by adding new feature selection methods and metrics.

Employing the widely used *sequential forward selection* technique, the AHFS algorithm incrementally builds the selected feature set by evaluating each candidate feature provided by the measures at each iteration. This adaptability enables the algorithm to explore various feature selection methods and evaluation measures, resulting in a comprehensive and effective approach.

To address the lack of a universally optimal feature selection algorithm, the AHFS algorithm uses the applied learning method as an independent evaluation tool. By constructing candidate model configurations and selecting features based on the specific learning model used, it ensures adaptive feature selection tailored to the learning model's needs. The algorithm produces feature order, accuracy, model error and concrete models as outcomes, contributing to a thorough evaluation of model performance.

Feature analysis through graphical representation

The feature selection process results were visualised and analysed using cluster and trend analysis techniques. Graphical representations provided insights into the characteristics and relationships of the selected features, identifying patterns, trends or anomalies. The effectiveness and suitability of these features for their intended purpose were assessed, evaluating their importance in capturing the underlying data characteristics and their impact on subsequent analysis or modelling tasks. The findings were contextualised within the broader research objectives and existing literature, highlighting the significance of specific features.

The graph-based method allows the database to be adjusted as needed. Since corruption is complex and often imperceptible, no single model can predict its level accurately. Each model's results should be examined carefully, noting any trends or differences. Diagrams should be organised based on social, economic and other factors in relation to perceived corruption (CPI). Function analysis measures the interaction between corruption (CPI) and other factors, with corruption acting as a disruptive force in society (Lambsdorff 2008).

The diagrams illustrate connections and trends that help determine the situation of a country or region. By analysing the distance from the axes and trends, three important factors can be identified. The charts classify according to the value of other factors and the CPI provided by TI, with the third factor relating to the statistical trend. The analysis results can be summarised as follows.

Results

Household out-of-pocket expenses

Out-of-pocket costs are one element of the Bertelsmann Institute's *'Health'* index. The indicator describes how much of the healthcare costs households themselves must cover. This is measured by a ratio (%):

healthcare costs paid by households/total healthcare expenditure (state + households).

This percentage is on the x-axis of our graph. Although the data are quite scattered, a 'quasi' trend can be drawn (see in Appendix Figure A1).

On the upper left are the northern European states, New Zealand and Luxembourg, and on the lower right is Mexico. If we take this trend as a basis, then it can be said that the more medical treatment citizens have to pay for – i.e. the less the state supports – the higher the perceived level of corruption. This is logical, because the more services and help you have to pay for out of pocket, the more we feel that this is a solution in other areas of life as well: I often have to pay someone if I want my case to be taken care of. And this is a kind of corruption.

The other ('quasi') trend: below a CPI score of y = 60, there are typically Middle Eastern and Southern European countries, as well as Turkey (which is practically also part of Southern Europe), South Korea and Mexico, extending all the way to the lowest and among the highest public health subsidies (x-axis). Of these, the healthcare of Croatia, the Czech Republic and Slovenia is the most subsidised by the state (only about 10% of the citizens' deductibles), while the healthcare of Cyprus and Latvia is the most expensive for citizens. This trend indicates that in Central, Eastern and Southern Europe, the perceived level of corruption is independent of the extent of 'reach into the pocket', i.e. the health expenditure of households. This seems to be confirmed by the trend that the level of perception of corruption in northern European countries among healthcare systems receiving high state support is much lower than in former socialist states with the same proportion of support, as well as the fact that there is no actual trend on the graph – the data points are completely scattered.

Infant mortality

The Bertelsmann SGI indicator measures the infant mortality rate per 1,000 live births. The lower this ratio, the more favourable it is; therefore, the higher the score in the SGI. The graph shows the ratio of deaths per thousand live births (see in Appendix Figure A2).

Although a well-explained trend line cannot be fitted to the points, it can be said that increasing infant mortality is associated with a less favourable image of corruption. The performance of Mexico and Turkey is also the most unfavourable in this indicator. A CPI score below 50 is similarly unfavourable in Romania and Bulgaria, which rank 3rd and 4th in the infant mortality ranking respectively. Infant mortality is one of the most common indicators of health conditions, so it can be evaluated together with the previously evaluated out-of-pocket expenses indicator, and thus the health situation and the perception of corruption can be examined more thoroughly.

Female labour force participation

The female labour force participation rate is measured by the Bertelsmann SGI index by the ratio of women and men present in the labour market. Its optimal value is 1. In the examined countries, the ratio varies between 0.48 (Turkey) and 0.97 (Lithuania). Turkey and Mexico are also at the bottom of the ranking in this indicator, and the Scandinavian and Baltic states perform best (as well as Israel). With the same labour market rate, the CPI spreads over a significant range: with a rate of 0.75, the CPI score is between 43 and 76, while with a rate of 0.95, it is between 57 and 91 (see in Appendix Figure A3).

The trend of each country shows a different course. With a similar labour market ratio, the CPI score is increasing in the southern European states: Greece, Bulgaria, Italy, Spain, Portugal, Cyprus and Malta. In the developed Asian economies of Japan and South Korea, the increasing participation of women in the labour market is accompanied by a decreasing CPI score. This result can be explained by the traditional Asian cultural view. According to this, women are responsible for giving birth and managing the household, while corporate work and especially decision-making are for men. The increase in the proportion of women in the labour market in these countries is not caused by a cultural change but by a growing labour shortage due to an ageing society. This problem is so urgent that it can no longer be solved by increasing the birth rate, as this will mean the labour force in about twenty years.

Now women must go to work to sustain the struggling economy (Xu 2023). The time series trend of the labour market participation rate of South Korean women is the same as the Japanese data. The reasons are also similar (Jang 2023). An interesting result is that Turkey's trend is completely similar to Korea's and Japan's. In this case, the role of the religious background is decisive. With unchanged female labour market participation, the CPI is decreasing in the Scandinavian and Nordic states (Denmark, Norway, Sweden, Finland, Iceland) and the Visegrad countries.

The only thing that can be said about the relationship between the indicator and the CPI is that an increase in the proportion of women in the labour market is associated with a higher CPI, i.e. it indicates a more favourable perception of corruption.

Therefore, no trend can be detected between the female participation rate and the CPI either. At the same time, in the band between the two lines, the tendency can be felt that an increase in the proportion of women in the labour market is associated with a higher CPI, i.e. it indicates a more favourable perception of corruption. However, the significant dispersion of the data in the band shown in Appendix Figure A3 allows only the above statement; the indicator is not suitable for judging and predicting the perception of corruption. However, the level of corruption perception of northern and western countries is higher than that of southern and eastern countries (the dividing line is close to the CPI score of 60).

Total researchers

The total researchers indicator is measured by SGI as the number of researchers per 1,000 people. The index is a sub-index of *research, innovation and infrastructure* (see in Appendix Figure A4).

As an approximate trend, it can be observed that an increase in the share of researchers among employees (a positive change) is associated with a higher CPI score (which is also favourable). Among the out-of-trend data, the level of corruption perception is better than expected based on the frequency of researchers in Chile (researchers ~1/1,000; CPI score ~70). Since Chile's performance has proven to be more favourable than the trend in several cases, the authors saw it as worthwhile to look for the reason for this. The South American country performs surprisingly well in the CPI score despite the fact that the region is traditionally classified as one of the world's most corrupt regions. The CPI score indicates that the institutional trap is not here for some reason. In 2023, Chile was 29th out of the 180 countries examined in the CPI ranking. The good performance can be explained by several factors. Legislation allows citizens to request public data and the operation of the police (Carabineros) is transparent and legitimate (Rolando 2017).

Since Mexico only achieved a CPI of around 30 with the same proportion of researchers, it can be concluded that the improvement of the CPI with the increase in the frequency of researchers cannot be generalised as a law. The same is confirmed by the example of Israel, which deviates from the trend in an unfavourable direction (researchers ~ 12.5 → 16.5/1,000 employees; CPI score ~ 64 → 59). In addition to the unchanged frequency of researchers in this country, the perception of corruption worsened by 5 CPI points in the period between 2016 and 2021. Like Israel, Greece (researchers ~ 6.8 → 11.6/1,000 employees; CPI score ~ 43 → 52) and also South Korea (researchers ~ 12.6 → 16.6/1,000 employees; CPI score ~ 55 → 63). However, with the increase in the number of researchers in these countries, the perceived corruption also changes in a favourable direction.

Biocapacity

Corruption can also have a negative impact on the state of the environment, for example by circumventing legislation and issuing permits without adequate protection (Xu et al. 2024). Sustainability is measured by several metrics, one of which is biocapacity, which is the area of agricultural land per person (Niccolucci et al. 2012). On the other hand, biocapacity is the Earth's ability to neutralise and renew the load caused by human activity (Pathak 2020). The SGI uses the latter interpretation but uses the hectare/capita value as the unit of measurement (see in Appendix Figure A5).

Regarding the biocapacity–CPI relationship, it can be said that the vast majority of the examined countries, regardless of the level of perceived corruption, belong to the 0–5 ha/capita biocapacity range. This clearly indicates the finite capacity of our planet and shows that the extent of biocapacity is given, since the land area cannot be increased. However, the groups G1, G2 and G3 fall outside this regularity. The G1 group includes the economically developed countries with low population density: Australia, New Zealand, Canada (population density < 20 capita/km², GDP > 1 billion USD/y), as well as the rich and sparsely populated Scandinavian countries of Europe. Population density seems to play an important role. In fact, population density can be considered the reciprocal of biocapacity. The same phenomenon applies to Iceland, the only member of the G2. Part of these countries was analysed by Bartha and his research group (2013). It was found that Norway has already reached the limit of its biocapacity, while Sweden has considerable room for manoeuvre. In their work, the OECD member states were classified into clusters based on their future opportunities, external and internal potential. The possibilities of the future are determined by the development of human capital and education, as well as the social responsibility of companies. The external potential is influenced by the decisions of the country brand and credit rating agencies. Internal opportunities are based on the competitiveness of companies, government efficiency and the availability of resources. In terms of this indicator, the best-performing cluster includes Australia, Canada, Norway, Finland and Sweden highlighted in the G1 group, and a separate cluster is formed by the G2 group country, Iceland, which performs well in the area of future and internal potential. And New Zealand moved down by one cluster, mainly due to its weaker internal opportunities. New Zealand has already been mentioned several times during this research, so the authors examined it in more detail. Since, according to Bartha et al. (2013), it was not among the top performers primarily due to internal opportunities, the investigation also focused on this question. In 2023, the country ranked 14th in the recognised country brand ranking (Ipsos 2023), behind the countries of the G1 group. The country also fell back in the CPI ranking. Yet building the country's 'Brand New Zealand' brand required significant resources. This country brand became the basis of very significant income from exports and tourism. However, the signs of a decrease in government efficiency are becoming more and more perceptible. This is shown by the misuse of subsidies during the pandemic and the weakening supervision of the spending of public funds (Raskovic 2024). These factors may contribute to the fact that New Zealand is still in the G1 group, but its position is deteriorating. This is confirmed by other indicators.

The situation of the Baltic states is quite different. In 2002, Taagepera reported on the still strong post-Soviet legacy and the need for change. Eight years later, Wadsworth et al. (2010) write about a low level of corruption in all three countries. Rogulis (2023) already emphasises Estonia's best performance. The reason for this, in his opinion, is stable and efficient governance and citizen self-awareness. These lead to a low level of corruption, which increases citizens' trust in state institutions. The direction of development is therefore exactly the opposite of that in the case of New Zealand.

Homicide rate

This index has one component (national homicide rate per 100,000 people – hard data). The index scores between 1 and 10 (1 = highest rate, 10 = lowest homicide rate). The graph shows the rate per 100,000 people on the x-axis. It is important to know that, according to the literature, the causal relationship between homicides and corruption is in the opposite direction to the current investigation: the perceived level of corruption cannot be inferred from the number of homicides, but the evolution of the number of homicides can be predicted from the perceived level of corruption (Chainey et al. 2021). That is why all countries except Mexico are concentrated in the band between 5 and 6 points (see in Appendix Figure A6).

The relationship between corruption and crime is, of course, much more complex than the evolution of the perception of corruption being judged from the evolution of the homicide rate. Throughout its long history, corruption has been linked to other crimes and organised crime at many points (Beare 1997, Chêne–Hodess 2008). There have been several studies on the interdependence of corruption and money laundering (Barone et al. 2022, Chaikin–Sharman 2009). One modern form of these acts is the use of cryptocurrencies and blockchain technology to make financial transactions more difficult to detect (Choo 2015). Another aspect of corruption is the involvement of politics. For a long time in history, nepotism, for example, was not considered corruption, as it was the form of inheritance of power in ruling dynasties. Criticism of the relationship between politics and corruption has appeared since the modern era (Engels 2016). The phenomenon still exists today (Hegewald–Schraff 2024, Hyde 2021, Stolton 2023, Wax et al. 2023). Even more harmful and dangerous than the above forms are the connection between organised crime and politics through corruption (Capitani 2007, Montalbano 1987, Schultz 2023, Sergi–South 2016). This often takes the form of grand corruption (IMF 2019) and affects public funds. Due to the undoubted economic impact, it may be worthwhile to examine corruption together with several crime indicators. There is already an example of this in the literature (Inzelt et al. 2014, Sebestyén 2005).

Conclusion

The authors' research examined the extent to which macroeconomic statistical indicators can predict the CPI. The first part of the study analysed those indicators that showed a clear trend in relation to the CPI through graphical analysis. This second part deals with indicators that do not show a well-defined graphical trend, but nevertheless enrich the system of relationships between corruption perceptions and macroeconomic indicators with essential information. The indicators examined cover various socio-economic areas – including healthcare, public safety and crime, the labour market, gender equality, scientific and technological development, and sustainability issues. The analysis revealed that none of the indicators alone is suitable for predicting CPI values. At the same time, it became clear that the perception of corruption is closely related to life situations that may seem distant from the concept of corruption at first glance.

One of the most important findings of the research is that the more citizens are forced to finance public services – such as healthcare – out of their own pockets, the higher the perceived level of corruption. This is illustrated by the relationship between co-payments and the CPI. However, indicators reflecting basic health conditions – such as infant mortality – did not show a strong correlation with the perception of corruption. One possible reason for this is that the countries studied belong to the economically developed category (OECD, EU), so infant mortality is relatively low. Further examination of this issue, especially in developing countries, may be the subject of future research. Another important result is that the increase in women's labour market participation has a positive effect on the perception of corruption – i.e. in countries where gender equality is better achieved, the perceived level of corruption is lower. This is also supported by the graphical analysis of the relationship between the female employment rate and the CPI.

Overall, it can be said that corruption is an extremely complex social and economic phenomenon, and to understand it, it is not enough to examine a single economic indicator. Although the research only covered developed economies, the high number of elements (41 countries) and the length of the period studied (2012–2021) allow general conclusions to be drawn. Given that the CPI is a subjective index based on expert opinions, the next stage of the research will examine the 'soft' indicators (expert assessments) of the Bertelsmann SGI database and analyse their relationship with the perception of corruption. This will provide an even more complex and nuanced picture of the social embeddedness of corruption.

Acknowledgement

The research reported in the paper received support by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory and AI EDIH Hungary (Project ID: 101120929). On behalf of Project *Testing Machine Learning Algorithms*, we thank for the usage of HUN-REN Cloud (https://science-cloud.hu/) that significantly helped us achieving the results published in this paper.

Appendix

Figure A1

Household pocket expenses and CPI

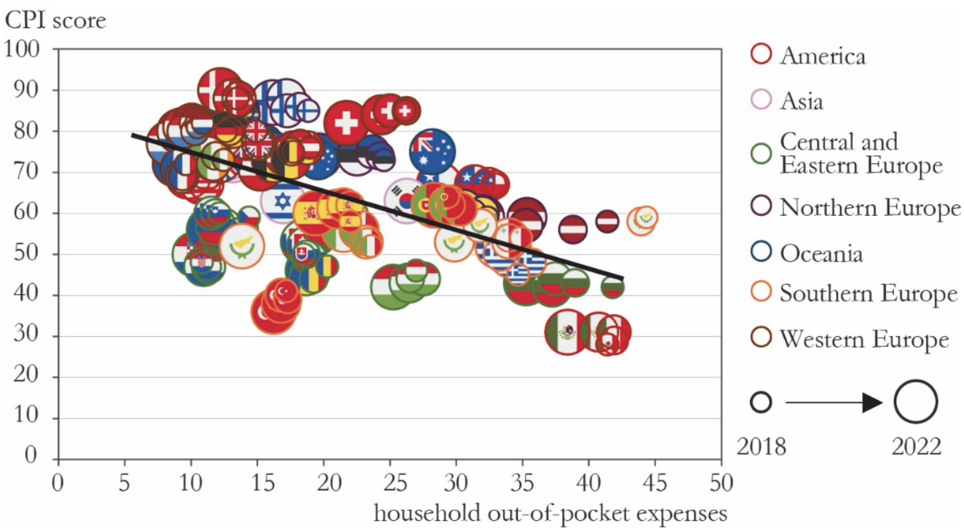


Figure A2

Infant mortality and CPI

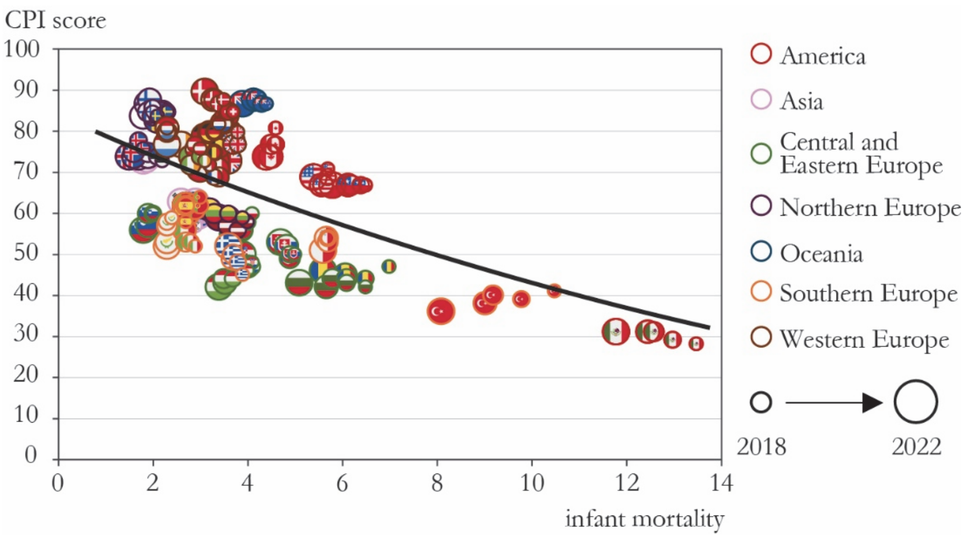


Figure A3

Female workforce and CPI

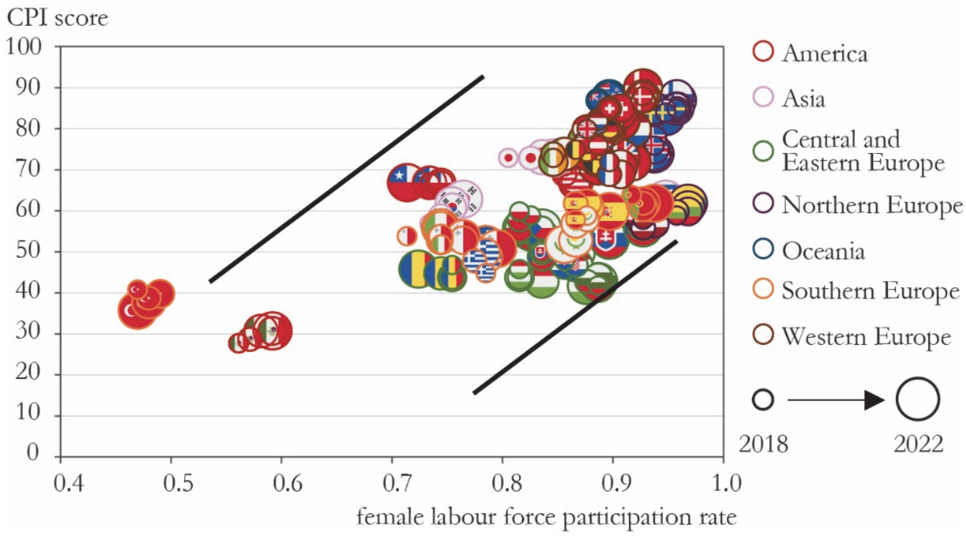


Figure A4

Researchers and CPI

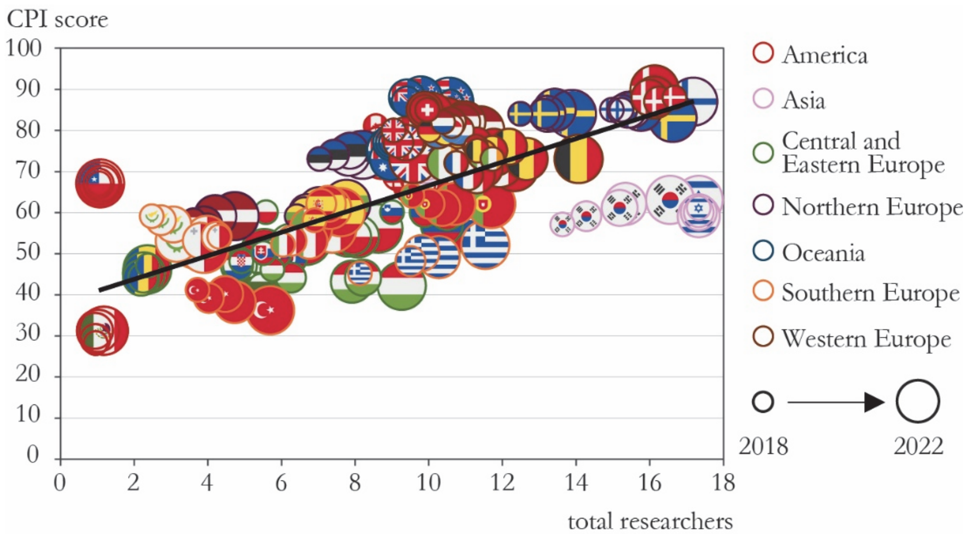


Figure A5

Biocapacity and CPI

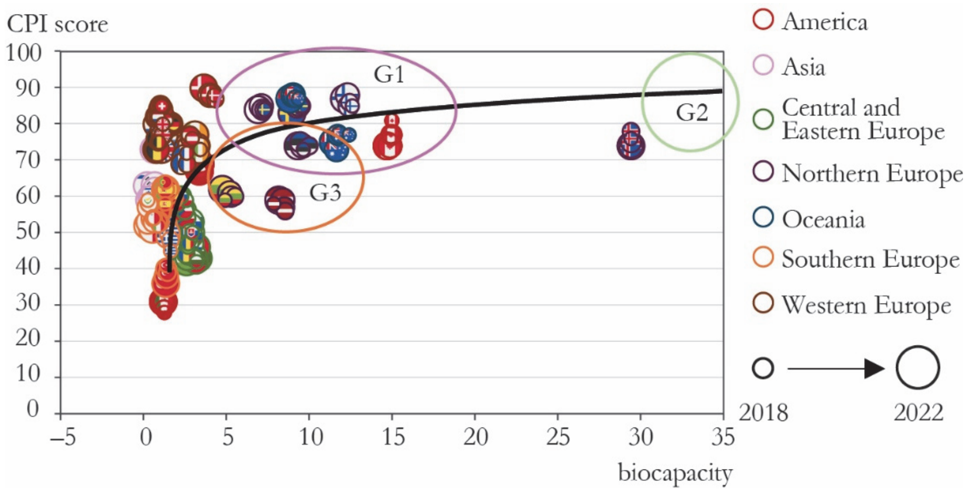
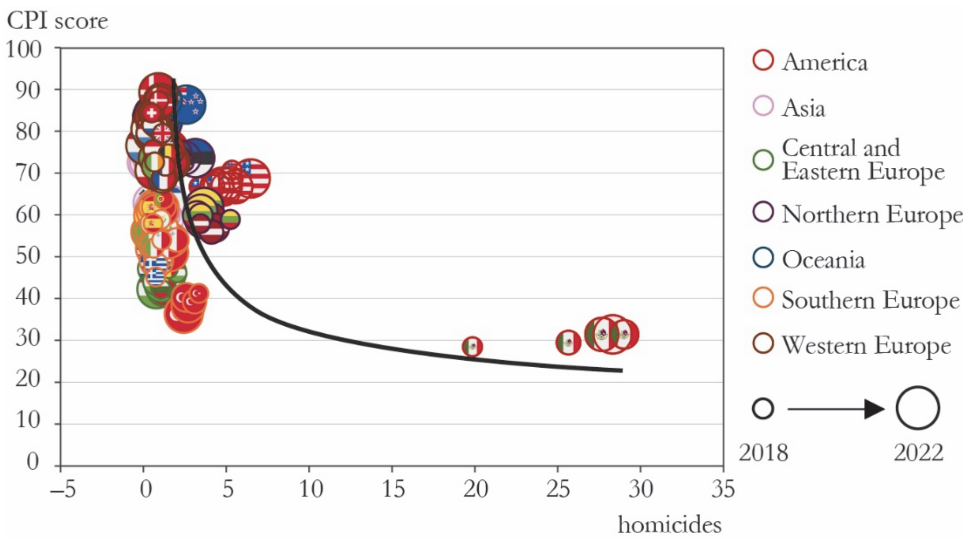


Figure A6

Homicide rate and CPI



REFERENCES

ABRAMO, C. W. (2008): How much do perceptions of corruption really tell us? *Economics (e-Journal)* 2 (1): 20080003. https://doi.org/10.5018/economics-ejournal.ja.2008-3

ATSIR, D. I.–SUNARYATI, S. (2018): The effect of corruption on international trade: a case study of Indonesian trade to nine countries *Eko-Regional Jurnal Pengembangan Ekonomi Wilayah* 13 (1): 19–26. https://doi.org/10.20884/1.erjpe.2018.13.1.1160

BARONE, R.–MASCIANDARO, D.–SCHNEIDER, F. (2022): Corruption and money laundering: you scratch my back, i'll scratch yours *Metroeconomica* 73 (1): 318–342. https://doi.org/10.1111/meca.12365

BEARE, M. E. (1997): Corruption and organized crime: lessons from history *Crime, Law and Social Change* 28 (2): 155–172. https://doi.org/10.1023/A:1008220609526

BELIAEVA, T.–SHIROKOVA, G.–WALES, W.–GAFFOROVA, E. (2020): Benefiting from economic crisis? Strategic orientation effects, trade-offs, and configurations with resource availability on SME performance *International Entrepreneurship and Management Journal* 16 (1): 165–194. https://doi.org/10.1007/s11365-018-0499-2

BUDSARATRAGOON, P.–JITMANEEROJ, B. (2020): A critique on the corruption perceptions index: an interdisciplinary approach *Socio-Economic Planning Sciences* 70: 100768. https://doi.org/10.1016/j.seps.2019.100768

CAMPOS, J. E.–PRADHAN, S. (2007): *The many faces of corruption: tracking vulnerabilities at the sector level* World Bank Publications. https://books.google.hu/books?hl=hu&lr=&id=Wy-oNmjqq-QC&oi=fnd&pg=PA389&dq=corruption+and+money+laundering&ots=sg32Oam2SQ&sig=BV-V5Ivb7Y93rpA3K_Yh58Upvr0&redir_esc=y#v=onepage&q=corruption%20and%20money%20laundering&f=false

CAPITANI, G. (2007): *Il generale Dalla Chiesa* [Film]. Endemol Shine Italy, Mediavivere.

CHAIKIN, D.–SHARMAN, J. (2009): *Corruption and money laundering: a symbiotic relationship* Springer.

CHAINEY, S. P.–CROCI, G.–FORERO RODRIGUEZ, L. J. (2021): The influence of government effectiveness and corruption on the high levels of homicide in Latin America *Social Sciences* 10 (5): 172. https://doi.org/10.3390/socsci10050172

CHOO, K.-K. R. (2015): Cryptocurrency and virtual currency: corruption and money laundering/terrorism financing risks? In: CHUEN, D. L. K. (ed.): *Handbook of digital currency* pp. 283–307., Academic Press. https://doi.org/10.1016/B978-0-12-802117-0.00015-1

DOMASHOVA, J.–POLITOVA, A. (2021): The corruption perception index: analysis of dependence on socio-economic indicators *Procedia Computer Science* 190: 193–203. https://doi.org/10.1016/j.procs.2021.06.024

DONCHEV, D.–UJHELYI, G. (2014): What do corruption indices measure? *Economics & Politics* 26 (2): 309–331. https://doi.org/10.1111/ecpo.12037

ENGELS, J. I. (2016): *A korrupció története* Corvina Kiadó, Budapest.

FHIMA, F.–NOUIRA, R.–SEKKAT, K. (2023): How does corruption affect sustainable development? A threshold non-linear analysis *Economic Analysis and Policy* 78: 505–523. https://doi.org/10.1016/j.eap.2023.03.020

HEGEWALD, S.−SCHRAFF, D. (2024): Corruption and trust in the European Parliament: Quasi-experimental evidence from the Qatargate scandal *European Journal of Political Research* 1475-6765:12654. https://doi.org/10.1111/1475-6765.12654

HUYSMANS, J.−MARTENS, D.−BAESENS, B.−VANTHIENEN, J.−VAN GESTEL, T. (2006): Country corruption analysis with self-organizing maps and support vector machines. In: CHEN, H.–WANG, F-Y.–YANG, C-C.–ZENG, D.–CHAU, M.–CHANG, K. (eds.): *International workshop on intelligence and security informatics* pp: 103−114., Springer, Berlin–Heidelberg. https://doi.org/10.1007/11734628_13

HYDE, R. (2021): Von der Leyen admits to Covid-19 vaccine failures *The Lancet* 397 (10275): 655–656. https://doi.org/10.1016/S0140-6736(21)00428-1

INZELT, É.−KEREZSI, K.−LÉVAY, M. (2014): *Korrupciós bűncselekmények a büntető igazságszolgáltatás tükrében* ELTE Állam és Jogtudományi Kar, Budapest.

KYUNG, H.−NAM, J. S. (2023): Insider trading in news deserts *The Accounting Review* 98 (6): 299–325. https://doi.org/10.2308/TAR-2020-0799

LAHMIRI, S. (2016): Features selection, data mining and financial risk classification: a comparative study *Intelligent Systems in Accounting, Finance and Management* 23 (4): 265−275. https://doi.org/10.1002/isaf.1395

LIANG, D.–TSAI, C. F.–WU, H. T. (2015): The effect of feature selection on financial distress prediction *Knowledge-Based Systems* 73: 289–297.

NAGY, Á.–DOMBI, J.–FÜLEP, M. P.–RUDICS, E.–HOMPOTH, E. A.–SZABÓ, Z.–DÉR, A.–BÚZÁS, A.–VIHAROS Z., J.–HOANG, A. T.–MACZÁK, B.–VADAI, G.–GINGL Z.–LÁSZLÓ, S.–BILICKI, V.–SZENDI, I. (2023): The actigraphy-based identification of premorbid latent liability of schizophrenia and bipolar disorder *Sensors* 23 (2): 958. https://doi.org/10.3390/s23020958

NÉMETH, E.–JAKOVÁC, K.–VARGHA, B. (2021): Módszertani szempontok a hazai korrupciós helyzet tudományosan megalapozott mérésének kialakításhoz *Belügyi Szemle* 69 (11): 1963–1980. https://doi.org/10.38146/BSZ.2021.11.5

NÉMETH, E.–VARGHA B., T.–PÁLYI K., Á. (2019): The scientific reliability of international corruption rankings *Pénzügyi Szemle* 64 (3): 319–336. https://doi.org/10.35551/PSZ_2019_3_1

NICCOLUCCI, V.–TIEZZI, E.–PULSELLI, F. M.–CAPINERI, C. (2012): Biocapacity vs ecological footprint of world regions: a geopolitical interpretation *Ecological Indicators* 16: 23–30. https://doi.org/10.1016/j.ecolind.2011.09.002

PATHAK, S. (2020): Ecological footprints of war: an exploratory assessment of the long-term impact of violent conflicts on national biocapacity from 1962–2009. *Journal of Environmental Studies and Sciences* 10 (4): 380–393. https://doi.org/10.1007/s13412-020-00626-5

PENG, Y.–ALBUQUERQUE, P. H. M.–KIMURA, H.–SAAVEDRA, C. A. P. B. (2021): Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators *Machine Learning with Applications* 5. https://doi.org/10.1016/j.mlwa.2021.100060

RASKOVIC M. (2024): A slide in global corruption rankings is bad for 'Brand NZ' – what can the government do? *The Conversation.* https://theconversation.com/a-slide-in-global-corruption-rankings-is-bad-for-brand-nz-what-can-the-government-do-222995

RAVISANKAR, P.–RAVI, V.–RAO, G. R.–BOSE, I. (2011): Detection of financial statement fraud and feature selection using data mining techniques *Decision support systems* 50 (2): 491–500. https://doi.org/10.1016/j.dss.2010.11.006

ROGULIS, D. (2023): Willingness to fight for one's own country: new variables to look at the Estonian leadership among the Baltic states *Lithuanian Annual Strategic Review* 20 (1): 105–118. https://doi.org/10.47459/lasr.2023.20.5

ROLANDO, A. I. (2017): How can we explain the lower levels of corruption perception in Chile and Uruguay? *Córdoba Global.* https://cbaglobal.com.ar/how-can-we-explain-the-lower-levels-of-corruption-perception-in-chile-and-uruguay/

SEBESTYÉN, T. (2005): Korrupciós trendek a magyar és nemzetközi adatok tükrében 1. rész *Polgári Szemle* 1 (3): 20–29. https://polgariszemle.hu/archivum/10-2005-aprilis-1-evfolyam-3-szam/20-korrupcios-trendek-a-magyar-es-nemzetkoezi-adatok-tuekreben-1-resz

SERGI, A.–SOUTH, N. (2016): 'Earth, water, air, and fire': environmental crimes, mafia power and political negligence in Calabria. In: ANTONOPOULOS, G. A. (ed.): *Illegal entrepreneurship, organized crime and social control: essays in honor of professor dick hobbs* pp. 5–100., Springer International Publishing. https://doi.org/10.1007/978-3-319-31608-6_6

TAAGEPERA, R. (2002): Baltic values and corruption in comparative context *Journal of Baltic Studies* 33 (3): 243–258. https://doi.org/10.1080/01629770200000081

TRANSPARENCY INTERNATIONAL (2024): *Corruption perceptions index 2012–2023.* https://www.transparency.org/en/cpi/2023

UDDIN, I.–RAHMAN, K. U. (2023): Impact of corruption, unemployment and inflation on economic growth evidence from developing countries *Quality & Quantity* 57 (3): 2759–2779. https://doi.org/10.1007/s11135-022-01481-y

UTOMO, V. G.–DEWI, R. A.–MARUTHO, D.–HIDAYAT, A. (2021): Analyze corporate anti-corruption disclosure with feature selection. In: *2021 International Seminar on Application for Technology of Information and Communication (iSemantic)* pp. 34–237., IEEE. https://doi.org/10.1109/iSemantic52711.2021.9573198

VIHAROS, Z. J.–KIS, K. B.–FODOR, Á.–BÜKI, M. I. (2021): Adaptive, hybrid feature selection (AHFS) *Pattern Recognition* 116. https://doi.org/10.1016/j.patcog.2021.107932

WADSWORTH, F.–SWARTZ, B.–WHEAT, J. (2010): Corruption in the Baltic state region *International Business & Economics Research Journal (IBER)* 9 (2). https://doi.org/10.19030/iber.v9i2.527

WALL, T. D.–MICHIE, J.–PATTERSON, M.–WOOD, S. J.–SHEEHAN, M.–CLEGG, C. W.–WEST, M. (2004): On the validity of subjective measures of company performance *Personnel Psychology* 57 (1): 95–118. https://doi.org/10.1111/j.1744-6570.2004.tb02485.x

WALTERS, R. (2013): Eco mafia and environmental crime. In: CARRINGTON, K.– BALL, M.– O'BRIEN, E.– TAURI, J. M. (eds.): *Crime, justice and social democracy: international perspectives* pp. 281–294., Palgrave Macmillan UK. https://doi.org/10.1057/9781137008695_19

WANG, W. K.–STEINBERG, M. I. (2010): *Insider trading* (3rd ed.) Oxford Press. https://repository.uclawsf.edu/faculty_books/18/

XU, R. (2023): *Japan's economy would gain with more women in science and technology* IMF. https://www.imf.org/en/News/Articles/2023/11/13/cf-japans-economy-would-gain-with-more-women-in-science-and-technology

XU, R.–PATA, U. K.–DAI, J. (2024): Sustainable growth through green electricity transition and environmental regulations: do risks associated with corruption and bureaucracy matter? *Politická Ekonomie* 10 (10): 1–23. https://doi.org/10.18267/j.polek.1420

INTERNET SOURCES

BARTHA, Z.–SÁFRÁNYNÉ GUBIK, A.–TÓTHNÉ S., K. (2013): Intézményi megoldások, fejlődési modellek University of Miskolc, Miskolc. https://mpra.ub.uni-muenchen.de/50901/1/MPRA_paper_50901.pdf (downloaded: January 2025)

BERTELSMANN STIFTUNG (2022): *SGI 2022 | Methodology*. https://www.sgi-network.org/2022/Methodology (downloaded: January 2025)

BERTELSMANN STIFTUNG (2023a): *SGI 2022 | Data*. https://www.sgi-network.org/2022/Data (downloaded: January 2025)

BERTELSMANN STIFTUNG (2023b): *SGI 2022 | Good Governance | Executive Capacity | Implementation | Regulatory Enforcement*. SGI 2024 | Governing with Foresight (downloaded: January 2025)

BUDIK, J.–EZR, V. (2015): *Corruption perceptions index, credit rating and global competitiveness of selected countries* VŠFS | Vysoká škola finanční a správní pp. 1–12. https://is.vsfs.cz/publication/5195/Budik_Ezr_Corruption_Perceptions_Index.pdf (downloaded: January 2025)

CARVALHO, R. S.–CARVALHO, R. N. (2016): *Bayesian models to assess risk of corruption of federal management units*. https://ceur-ws.org/Vol-1663/bmaw2016_paper_5.pdf (downloaded: January 2025)

CHÊNE, M.–HODESS, R. (2008): *Organised crime and corruption* U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute. https://www.u4.no/publications/organised-crime-and-corruption.pdf (downloaded: January 2025)

ELŐD, F. (2020): *Dél-Afrikával kerültünk egy szintre korrupcióban*. https://index.hu/gazdasag/2020/01/23/transparency_international_korrupcio_index_2020_korruptabb_lett_magyarorszag/ (downloaded: January 2025)

INTERNATIONAL MONETARY FUND [IMF] (2019): *Fiscal monitor: curbing corruption*. https://www.imf.org/en/Publications/FM/Issues/2019/03/18/fiscal-monitor-april-2019 (downloaded: January 2025)

IPSOS (2023): *The Anholt-Ipsos Nation Brands Index – 2023* Ipsos. https://www.ipsos.com/sites/default/files/ct/news/documents/2023-10/NBI_2023_Press_Release_Supplemental_Deck_WEB.pdf (downloaded: January 2025)

JANG, B. (2023): *Population decline and labour resources in Korea: current state and implications* Korea Capital Market Institute. https://www.kcmi.re.kr/common/downloadw?fid=26652&fgu=002001&fty=004003 (downloaded: January 2025)

LAMBSDORFF, J. G. (2008): *The methodology of the corruption perceptions index 2008*. Transparency International. https://files.transparencycdn.org/images/2008_CPI_LongMethodology_EN.pdf (downloaded: January 2025)

MONTALBANO, W. D. (1987): 19 get life terms in Sicilian mafia trial *Los Angeles Times*. https://www.latimes.com/archives/la-xpm-1987-12-17-mn-29590-story.html (downloaded: January 2025)

NAGY, G. (2018): *A korrupció fogalma mérése és hatása.* https://transparency.hu/wp-content/uploads/2018/02/A-korrupcio-fogalma-merese-%C3%A9s-hatasa.pdf (downloaded: January 2025)

PORTFOLIO.HU (2021): *Hármas holtversenyben az EU legkorruptabb országa lett Magyarország a Transparency International indexe szerint.* https://www.portfolio.hu/unios-forrasok/20210128/harmas-holtversenyben-az-eu-legkorruptabb-orszaga-lett-magyarorszag-a-transparency-international-indexe-szerint-467224 (downloaded: January 2025)

SAISANA, M.–SALTELLI, A. (2012): *Corruption perceptions index 2012: an updated methodology* European Commission Joint Research Centre. https://transparency.ee/files/files/lisad/2012_cpiupdatedmethodology_en.pdf (downloaded: January 2025)

SCHILLER, C.–HELLMANN, T.–PAULINI, P. (2023): *SGI codebook – sustainable governance indicators 2022* Bertelsmann Stiftung. https://www.sgi-network.org/docs/2022/basics/SGI2022_Codebook.pdf (downloaded: January 2025)

SCHULTZ A. (2023): Megmutatkozott a maffia hatalma az olasz bíróságon *24.hu*. https://24.hu/kulfold/2023/11/26/ndrangheta-olaszorszag-maffia-kalabria-forza-italia-luigi-mancuso-itelethozatal-drogkereskedelem/ (downloaded: January 2025)

SZEMÁN L. J. (2022): A korrupciós érzet magasabb, mint a valós korrupciós szint *Magyar Nemzet*. https://magyarnemzet.hu/belfold/2022/10/a-korrupcios-erzet-magasabb-mint-a-valos-korrupcios-szint (downloaded: January 2025)

STOLTON, S. (2023): New York Times sues EU over von der Leyen's Pfizer texts *Politico*. https://www.politico.eu/article/new-york-times-sue-european-union-ursula-von-der-leyen-pfizer-texts/ (downloaded: January 2025)

TANZI, V.–DAVOODI, H. (1997): *Corruption, public investment, and growth* International Monetary Fund. https://www.imf.org/external/pubs/ft/wp/wp97139.pdf (downloaded: January 2025)

WAX, E.–BRAUN, E.–WHEATON, S. (2023): 11 key Qatargate players – where are they now? *Politico*. https://www.politico.eu/article/qatargate-corruption-scandal-13-key-players-six-months-on-european-parliament/ (downloaded: January 2025)