Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Ivo Seitz, 28.04.2019
Social Data Science 2019

Google Searches Indicate Financial Crime in Countries

A theory of amplification of public interest by financial crime in countries and the crucial role of internet search in today's information gathering process, as well as perceived online anonymity, suggest the use of online interest measured by internet search volumes as indicator of financial crime in countries. This report finds significant correlation between Google search volumes for "money laundering" and "corruption" in countries and corresponding indicators. Thus, Google search volumes are proposed as complementary source for country risk assessments in financial crime prevention with specified limitations.

Review of Research Leveraging Google Trends Data

The popularity of Google searches reflects the relevancy of topics for a society as shown in an increasing number of studies [1]. As of 2018, over 650 research papers use Google Trends in areas such as IT, communications, biology, and economics [1]. At least 46 studies found Google searches to correlate with macroeconomic variables such as GDP, unemployment rates, prices or sales over time [2]. Moreover, Google searches can predict the activity of some infectious diseases much earlier than official reports [3] [4]. However, most research concerns time series and not necessarily cross-country analysis such as [5] where Google Trends data serves as indicator of online interest across countries and as such as a measure of society's concern with a topic. This report examines this research direction for money laundering and corruption.

Application of the Method

Risk assessments of countries with respect to ML and corruption suffer from a lack of accessible and reliable data allowing a complete comparison of countries. Moreover, difficulties arise due to differences in legal practice, attention to the topic, and data collection methods.

The current industry standard in financial crime prevention is based, besides internal assessments, to a large extent on few established and publicly available indicators [6] which partially rely on expert judgement and undergo periodic review resulting in changes over time. Contradictions between indicators occur. Moreover, the publisher of an indicator, often a supranational body, might be biased by conflict of interest, e.g. treatment of its member countries. Consequently, binary decisions, whether a specific country should be classified as sensitive or not, are a difficult but very relevant problem in financial crime prevention. Finally, data allowing for a critical challenge of indicators are rare or do not reflect the entire spectrum of specific financial crime risk.

Google search volumes, as presented in this report, could not only challenge but complement these mostly subjective sources. This would enable extensions of risk assessments to more granular analysis in the time or regional (sub-country) dimension. However, risk assessments should not only be based on Google search volumes due to misclassifications that occur with this method, its limited representativity, and other limitations presented in this report.

Research Hypotheses

The hypotheses of positive association are examined, specifically that *risk for money laundering (ML) / corruption is high (low) in countries with large (small) Google search volume for money laundering / corruption as rated by corresp. indicators*. The examined indicators are standard in financial crime prevention industry and are listed in the appendix along data source specifications.

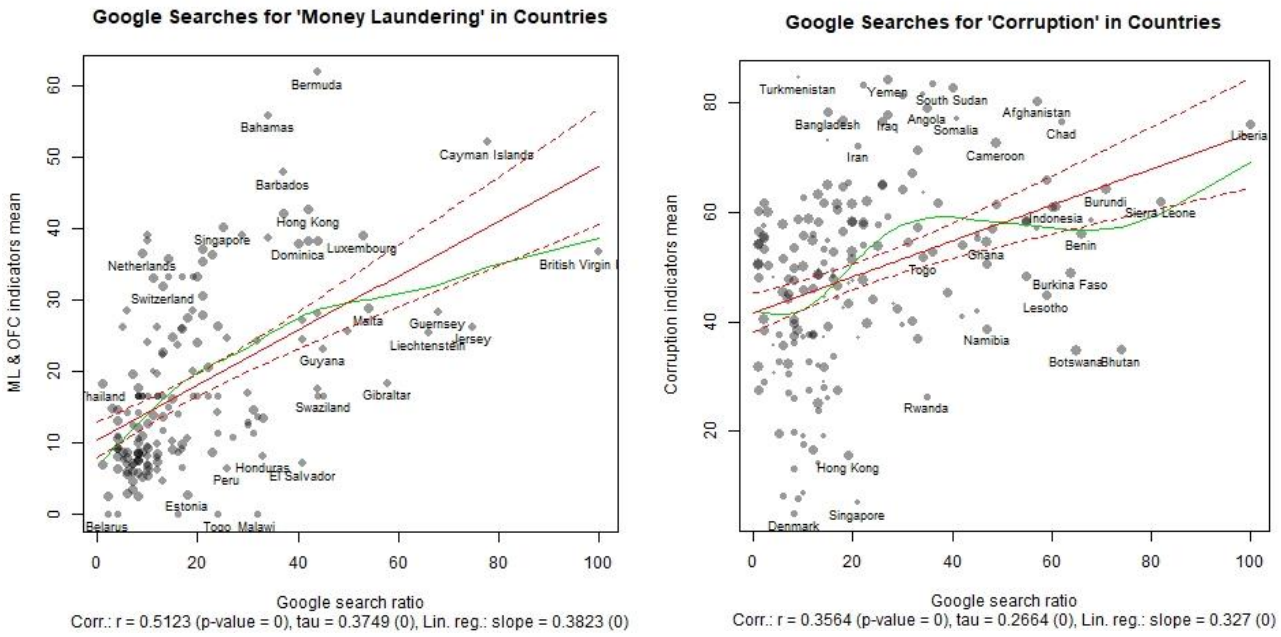


Fig. 1: Comparison between Google search topic *ML* / *Corruption* and mean of ML / corruption indicators. Each dot represents a country with a relative Google search volume (x-axis) and a mean of all available indicators (y-axis). The dot size relates to the number of available indicators. For detailed results refer to table and appendix. The green line shows a cubic smoothing spline fitted to the data and the red line a linear regression with 95 % confidence intervals and slope with p-value (null-hypothesis: no linear dependence) as indicated below the graph. Also printed below the graph are computed Pearson (r) and Kendall (tau) correlation coefficients with corresp. p-values (null-hypothesis: no positive correlation, one-sided test).

Results

For both money laundering and corruption, it is found that the countries with highest Google search volume are consistently rated as risky by at least some indicators: The nine countries with top Google search volume (≥ 50) for ML are offshore island states (British Virgin Islands, Cayman Islands, Jersey, Guernsey, Isle of Man) or small European territories (Liechtenstein, Gibraltar, Malta, Luxembourg). These nine countries are all on the FSF IMF 2000 list, six on IMF 2007, three on IMF 2018 list, and they have medium to high FSI Secrecy Score.

The eleven countries with most corruption searches (≥ 60), however, are besides Bhutan, Solomon Islands, and Indonesia, consistently African countries (Liberia, Sierra Leone, Burundi, Benin, Botswana, Burkina Faso, Chad, Ethiopia). All of them have at least one high to very high corruption indicator score, which is very often CPI, often WGI. Eight of them have a rank in the top 30 of at least one, often multiple, indicators. Liberia, as an example, has besides high CPI, WHR, and TRACE scores, the 2nd highest GCB and 4th highest IC score while Burundi has the 8th highest TRACE, 10th highest CPI, and 12th highest WGI score.

To test for positive correlation (one-sided test) between Google search volume and indicators, Pearson's product-moment correlation is calculated due to its wide-spread use. However, the distribution of Google search volumes violates the normality assumption of the hypothesis test for Pearson's correlation and Kendall's rank correlation test is provided as appropriate alternative. The significance results are comparable. Indicators which are binary lists were instead tested for a difference in means with a one-sided two-sample Wilcoxon rank test.

Test results show for ML on level 0.05 significant correlation for all indicator scores (table 1, Basel AML Index only significant for Kendall correlation) and significantly higher Google search volumes for EU Blacklist, IMF 2007, and IMF 2018 lists (table 2). For corruption, on the other hand, all indicators (except GCB and WHR for Kendall's correlation) show significant correlation with Google search volumes (table 3).

Money Laundering Indicator	Pearson correlation		Kendall correlation	
	r (estimate)	p-value	τ (estimate)	p-value
Basel AML Index	0.1163	0.1001	*0.1306*	*0.0178*
FSI Secrecy Score	*0.4059*	*0*	*0.3598*	*0*
FSF IMF 2000	*0.5849*	*0*	*0.4467*	*0*
mean (incl. binary lists)	*0.5123*	*0*	*0.3749*	*0*

Table 1: Correlation between Google search topic ML and ML indicator scores.

ML indicator	Wilcoxon p-value
FATF	0.2901
Egmont Gr.	0.1823
EU Blacklist	*0.0021*
IMF 2007	*0*
IMF 2018	*0.0016*

Table 2: Test results on higher Google search volumes for topic ML in countries on binary ML indicator lists.

To assess overall correlation, positive correlation between Google search volumes and indicator means is tested and found significant for both money laundering (r = 0.5123) and corruption (r = 0.3564). Country plots (fig. 1) show the overall positive dependences. For ML, there are neither countries which Google search volumes would wrongly predict as high-risk, nor such as low-risk countries but in the case of corruption, some high-risk countries (e.g. Turkmenistan) show low search volumes. However, these risk ratings are often based on only few indicators and might be inaccurate.

Corruption Indicator	Pearson correlation		Kendall correlation	
	r (est.)	p-value	τ (est.)	p-value
CPI	*0.3556*	*0*	*0.2790*	*0*
WGI	*0.3705*	*0.0002*	*0.3108*	*0*
GCB	*0.1591*	*0.0500*	0.0730	0.1362
IC	*0.3568*	*0*	*0.2993*	*0*
WHR	*0.1550*	*0.0259*	0.0611	0.1299
TRACE	*0.3718*	*0*	*0.2819*	*0*
mean	*0.3564*	*0*	*0.2664*	*0*

Table 3: Correlation between Google search topic Corruption and corruption indicators.

Thus, while not capturing all risk, Google search volumes can give helpful insight into the risk of a country. In summary, correlation with various indicators are found to be significant which provides strong evidence for the research hypotheses considering the variability in indicators.

Plots for individual indicators can be found in the appendix. Similar but less pronounced results can be found for search terms (in multiple languages) instead of search topics and are given in the appendix.

Theoretical Rationale

The results indicate that public interest in financial crime reflects its risk on the level of countries which could be explained by following local amplification process: Interest in financial crime as captured by information need does not only directly reflect criminal intent to actively engage in it or curiosity about doing so, but is also a secondary reflection of such crimes by spread of stories in social groups or the media. In general, the relevancy of a financial crime topic for a society is rooted in the problematic occurrence of financial crime and causes increased interest in it.

The decrease of public interest in stories with geographical distance to the story location (public interest as a local phenomenon) could allow its use as local indicator. Moreover, countries with large financial crime prevention efforts as conducted by governments and banks have higher information need on such topics but large efforts are a likely result of presence of actual financial crime. In all cases, stronger presence of financial crime in a country amplifies interest in such topics which suggests its use as a local indicator of true financial crime.

In accordance with this theory, the research hypotheses use online interest as reflection of public interest. This is encouraged by the ubiquitous, simple, and low-cost access to the internet as dominant in most of the modern world and its status as today's main resource for information. The use of the internet for inquiries on criminal topics could moreover be facilitated by perceived anonymity[1] and the theoretic local amplification process. Google is the most used online portal to search for information[2] – thus, Google searches provide a good approximation of online interest.

[1] The so-called *dark web* would allow for even more anonymity and is well-known for criminal uses. However, the conventional part of the internet provides simpler access and is wider spread.
[2] Google is by far the most used tool to gather information in almost all countries [7].

Limitations

First, it is assumed that the examined indicators reflect money laundering / corruption risk in countries. These indicators could be biased by the applied methods and subjective assessments by the creators. However, there is a lack of objective, comparable, and complete information for financial crime risk of countries and the indicators used in this report reflect the industry standard applied today in financial crime prevention. Additional inherent limitations of established indicators are not discussed here.

Second, online interest does not fully coincide with the appearance of financial crime, but a theoretical connection is presented in this report.

Third, there could be self-selection bias based on representativity of internet users: Internet populations of countries do not fully represent the entire population but could be biased towards people with technical affinity. Different countries could be differently represented on the internet.

Fourth, Google's preprocessing of Google Trends data is nontransparent, the declared steps are listed in the appendix. Also, it is not declared how topics of Google searches are determined, e.g. what searches count for a search topic. Moreover, Google Trends data is not showing for very low search frequencies. This leads to the exclusion of countries with low internet activity. These countries are excluded in this report (72 for topic *Money Laundering*, 51 for topic *Corruption*). Potentially, not all Google searches are considered (other terms, misspellings, other languages etc.) or some unrelated Google searches are considered by mistake (e.g. laundry).

Fifth, there could be biased search volumes for countries where Google holds a small market share of the search engine business – notable examples are only China and Russia and countries affiliated to those [7]. However, Google Trends data gives proportional results (see appendix) and Google is by far the most popular search engine in almost all countries [7].

Future Research

Besides research on the causation of the discovered correlation, another suggestion for future research is a generalization to other search terms and indicators concerning e.g. terrorism financing, or in other areas than financial crime. Further, one could attempt to predict indicator changes in the time dimension. However, the rescaling of Google Trends data poses a difficulty in gathering data – data would need to be collected for each country individually and then aggregated. Moreover, instead of assessing online interest, one could assess the possibility of using online communication as proxy for financial crime. The geotag of tweets would allow a similar analysis (or tweets including e.g. *money laundering [Country]* for each country). Another alternative are news such as Reuters articles or Google news searches (also available on Google Trends). Sentiment analysis of such texts could provide additional insight.

Conclusion

The results largely support the hypotheses that public interest in financial crime topics within countries are positively associated with the financial crime risk in these countries. Consequently, the application of Google search volumes as complementary variables of financial crime risk is suggested for financial crime prevention with given limitations. The observations could be explained by a sociological theory of amplification of public interest, the central role of internet search in today's information gathering process, and perceived online anonymity.

Appendix

I. Money Laundering Indicators

- Financial Action Task Force (FATF) high-risk and other monitored jurisdictions (expert-based)

Financial Action Task Force (FATF) identifies and publishes jurisdictions with strategic anti-money laundering (AML) or counter-terrorism financing (CFT) deficiencies three times a year. The basis is a review by the International Co-operation Review Group (ICRG). This list includes 13 countries as of April 2019.
Source: http://www.fatf-gafi.org/countries/#high-risk (27.04.19)

- 2018 Basel AML Index (composite)

The International Centre for Asset Recovery, part of the Basel Institute on Governance, publishes annually the Basel Anti-Money Laundering Index which is an independent ranking assessing the risk of money laundering and terrorist financing. It is a composite index based on public sources and third-party assessments. The version of 2018 includes 129 countries. This report min-max scales the 2018 overall scores such that the values are between 0 (lowest risk of money laundering) and 100 (highest risk of money laundering).
Source: https://index.baselgovernance.org/ (27.04.19)

- Egmont Group Members (network of Financial Intelligence Units)

The Egmont Group of Financial Intelligence Units (FIUs) is a 1995 founded network of 159 FIUs which are central, national agencies with the mission to collect and analyze information on financial activities suspected of being money laundering or terrorism financing and to inform public prosecution agencies if sufficient evidence is found. The Egmont Group has the purpose of providing a global forum for FIUs to improve cooperation and provides various support to its members. A country which has no FIU being Egmont Group member is viewed as non-cooperative and might not show sufficient efforts in combatting money laundering.
Source: https://en.wikipedia.org/wiki/Egmont_Group_of_Financial_Intelligence_Units (27.04.19)

- Offshore Financial Centres: FSF-IMF 2000 (qualitative), IMF 2007 (quantitative), IMF 2018 (quantitative)

A country is an Offshore Financial Centre (OFC) if the largest users of the financial sector are nonresidents. Financial Stability Forum (FSF) and International Monetary Fund (IMF) identified 48 countries (updated) as OFCs (grouped into 3 groups according to quality of supervision and cooperation in the country, rated in this report as 33, 66, and 100; countries not on the list as 0) based on a qualitative approach in 2000. Using a quantitative approach, IMF published a revised list of OFCs (without grouping) in 2007 (22 countries) and 2018 (8 countries). The classification as an OFC does not necessarily coincide with higher money laundering rates but successful OFCs facilitate tax evasion and money laundering [10].
Source: https://en.wikipedia.org/wiki/Offshore_financial_centre (27.04.19)

- Financial Secrecy Index (FSI) (composite)

Tax Justice Network ranks countries according to secrecy in the financial sector based on 20 indicators. Similarly to OFCs, financial secrecy does not directly imply higher money laundering rates but can serve as an indicator. This report min-max scales the 2018 secrecy score such that the values are between 0 (least secrecy) and 100 (most secrecy).
Source: https://www.financialsecrecyindex.com/introduction/fsi-2018-results (27.04.19)

- EU list of non-cooperative tax jurisdictions (EU Blacklist)

The member states of the European Union (EU) agreed on 05.12.18 on a list of non-cooperative tax jurisdictions and updated it subsequently. The list is based on screenings and dialogues with non-EU countries, to assess them against criteria for good governance such as tax transparency, fair taxation, the implementation of OECD BEPS measures and substance requirements for zero-tax countries. This report considers the version of 12.03.19 including 15 countries.
Source: https://ec.europa.eu/taxation_customs/tax-common-eu-list_en (27.04.19)

II. Corruption Indicators

[9] reviews several indicators of corruption and concludes "that the aggregate survey-based indicators of corruption, especially the Corruption Perceptions Index and the World Bank's Control of Corruption indicator […] are the most valid measures of the magnitude of overall corruption in many country contexts." [9]. However, cross-checking of indicators is recommended [9]. "while both objective and subjective indicators of corruption [are good indicators], the most appropriate indicators are the composite subjective indicators: the CPI [Corruption Perceptions Index] and the CC [Control of Corruption Governance Metric]" [9].

- Corruption Perceptions Index (CPI) (aggregate survey-based)

Transparency International releases this annual rating of the relative degree of corruption every year since 1995 and describes it as "the leading global indicator of public sector corruption". The version of 2018 is based on 13 surveys and expert assessments to measure public sector corruption in 180 countries and territories, giving each a score from zero (highly corrupt) to 100 (very clean). This report min-max scales the negative 2018 scores such that the values are between 0 (lowest perception of corruption) and 100 (highest perception of corruption).
Source: https://www.transparency.org/cpi2018 (25.04.19)

- Control of Corruption (WGI) (aggregate survey-based)

Control of Corruption is one of six dimensions in World Bank's Worldwide Governance Indicators (WGI) which is based on a large number of enterprise, citizen and expert surveys, in total over 30 data sources. The values (estimate) reflect a rating of governance which ranges from -2.5 (weak) to 2.5 (strong) governance performance). This report min-max scales the negative country means of available data in years from 1996 up to 2017 such that the values are between 0 (strongest governance) and 100 (weakest governance).
Source: http://info.worldbank.org/governance/wgi/#home (27.04.19)

- Global Corruption Barometer (GCB) (citizen survey)

Every few years, Transparency International (TI) conducts world-wide public opinion surveys asking citizens about their direct personal experience of corruption. TI states that this is the largest public opinion survey on corruption. The version of 2017 covers 119 countries, territories and regions and is based on interviews with 162,136 adults from March 2014 until January 2017.
Question 3: *Total bribery rates by country* (*excluding no contact:* percentage of survey participants who had contact with a government service in the past 12 months and had to pay a bribe, give a gift, or do a favor in order to receive the service) is assessed in this report as it reflects best experience of corruption.
Source:
https://www.transparency.org/news/feature/global_corruption_barometer_citizens_voices_from_around_the_world (25.04.19)

- World Bank Indicator *IC.FRM.BRIB.ZS* bribery incidence (IC) (enterprise survey)

This bribery incidence is based on World Bank enterprise surveys and represents the percentage of firms experiencing at least one bribe payment request across 6 public transactions dealing with utilities access, permits, licenses, and taxes. 150 to 1800 (depending on size of economy) Business owners or top managers are surveyed every year. This report min-max scales the country means of all available data in years from 2006 up to 2018 such that the values are between 0 (fewest bribery incidences) and 100 (most bribery incidences).
Source: Helliwell, J., Layard, R., & Sachs, J. (2019). World Happiness Report 2019, New York: Sustainable Development Solutions Network, https://data.worldbank.org/indicator/IC.FRM.BRIB.ZS (27.04.19)

- World Happiness Report (WHR) Perception of Corruption (citizen survey)

The World Happiness Report is an annual survey about multiple topics of perceived happiness in 156 countries and is produced by the United Nations Sustainable Development Solutions Network.

This report considers *Corruption Perception* which is the national average of the survey responses to two yes/no questions in *Gallup World Poll*: “Is corruption widespread throughout the government or not?” and “Is corruption widespread within businesses or not?”
This report min-max scales the country means of all available data in years from 2005 up to 2018 such that the values are between 0 (lowest perception of corruption) and 100 (highest perception of corruption).
Source: https://worldhappiness.report/ed/2019/ (27.04.19)

- TRACE Bribery Risk Matrix® 2018 (TRACE) (aggregated survey-based)

TRACE published annually bribery risks of 200 countries which is a combination of the four domains *Business Interactions with Government, Anti-bribery Deterrence and Enforcement, Government and Civil Service Transparency* and *Capacity for Civil Society Oversight*. This report uses the 2018 total score which ranges from 0 (lowest bribery risk) to 100 (highest bribery risk).
Source: https://www.traceinternational.org/trace-matrix (27.04.2019)

III. Google Trends: Data Source

Google is the most-used web search engine in the world with billions of searches daily. Almost everywhere a Google search is the first step to receive information on just any topic. This report examines Google search volume data from Google Trends, https://www.google.com/trends (28.04.2019). The range is world-wide and the time period is from 01.01.2004 to 28.04.2019.

Google provides relative popularity of search volumes, i.e. the number of searches for *Money Laundering* or related searches in a given country for a given time period is divided by the total number of searches in the country and time range. This allows to compare search volumes between countries (and time ranges) without introducing a bias for countries with larger internet populations than other countries. The relative search volumes are then scaled on a range of 0 to 100 such that the country with the highest relative search volume receives the value 100. The value 0 (or blank) represents low search volumes, not the absence of any searches related to *Money laundering*.

One can either use search terms or topics to query Google Trends. The difference and handling in this report is explained in a following section.

There is also an option to include low search volumes. Analysis shows that this includes some countries which have otherwise value 0 but other countries still show value 0. This report chooses this option to include low search volumes in order to allow the analysis of countries with small internet populations which are of particular interest for financial crime risk.

Note: The ratio between values with and without this option should theoretically be consistent (rescaling. However, they show small differences (Belgium has value 5 with option, value 8 without option but Croatia has value 8 with option, value 19 without option). This is not explained by a rounding of the values to integers alone. This suggests that either multiple rounding steps are applied in data preprocessing or Google adds some random noise to the data for anonymization. However, the discrepancy seems to be at most 2 value points and is neglected in the analysis.

Google does not declare the exact preprocessing of data but simply states: “Google Trends data is an unbiased [random] sample of Google search data. Only a percentage of searches are used to compile Trends data.”[3] Moreover, duplicate searches from the same person over a short time period, as well as searches with special characters such as apostrophes, are removed.

Source:
https://support.google.com/trends/answer/4365533?hl=en (05.03.2019)

[3] https://support.google.com/trends/answer/4365533?hl=en (05.03.2019)

IV. Google Trends: Difference between search terms and topics

Search terms show matches for all terms as specified, i.e. mainly in the language given. If there are multiple words (separated by whitespace only), searches including the words in any order with other words at any position will be considered. *money laundering* as search term will therefore have a bias for English speaking countries and include searches such as *money to pay for laundering*. Quotation marks restrict the considered searches to those including the specified words in the specified order without other words in-between. However, the search term *"money laundering"* will still have a bias for English-speaking countries. Different languages could be added by a + (to be understood as logical *or*) such as in *"money laundering"* + *"lavado de dinero"*. However, this would require a focus on certain language and introduce a bias for the countries speaking these languages.
On the other hand, a topic is a group of terms that share the same concept in any language. The topic *Money Laundering* will include searches in multiple languages and related searches like *structuring*. In order to consider all possible languages and include related searches, this report analyzes the topic *Money Laundering*. This has also the benefit that larger search volumes can be analyzed which are especially beneficial for countries with low search volumes such as small island countries which are of interest for this report. Refer to the comparison of the search term *money laundering* and the topic *Money Laundering* at https://trends.google.com/trends/explore?q=%2Fm%2F04xh9,money%20laundering (05.03.19)

The topic *Money Laundering* shows more correlation with established indicators than search terms *"money laundering"* and the situation is the same for corruption. A plausible explanation is that the topic includes related searches in more languages which allows more countries to show non-negligible search volumes and can result in higher accuracy in terms of a more flexible connection between Google searches and interest in money laundering / corruption. Thus, topic is the primary source for Google search volume in this report. However, also search terms are assessed for a complete analysis and the more involved derivation of search terms is explained in the following.

Selection of search terms:

In contrast to a topic, a search term is not automatically translated into other languages. Hence, to avoid a bias towards English-speaking countries by the use of the search term *"money laundering"* alone, the expression was translated into multiple languages and the quoted expressions were concatenated by +[4].

Because the level of analysis are countries, the selection of languages is not simply based on most spoken languages but the languages with the highest number of countries recognizing this language as official language [8] were considered: English, French, Arabic, Spanish, Portuguese. Other languages have been assessed but no significant contribution was observed of any major language either in the top ten of native speakers or number of countries with this language as official language. Moreover, Google's market share in the search engine business was assessed. Results show that the countries where Google is not the most used search engine are mainly countries affiliated to Russia or China, but Russian and Mandarin are not part of the mentioned five languages.

Since Google Trends queries have a maximal length of 100 characters, a selection of expressions[5] had to be taken which was conducted as follows: Full expressions are used to ensure capturing precise searches about the topic (e.g. excluded colloquial abbreviation (French) *blanchiment* which could be related to laundry or bleaching of teeth). If multiple expressions are available in a language, the expression with the largest Google search volume is chosen (determined by the *compare* functionality of Google Trends). Following translations were considered:

[4] refer to appendix for an explanation of the effect of quotes and +
[5] refer to appendix for a list of considered translations

Language	Translation with largest Google search volume	Alternative translation
English	money laundering	–
French	blanchiment d'argent	blanchiment (de l'argent)
Arabic	غسيل الأموال	تبييض الأموال
Spanish	lavado de dinero	blanqueo de capitales
Portuguese	lavagem de dinheiro	branqueamento de dinheiro
(German)	Geldwäsche	–
(Mandarin)	洗钱	–

Language	Translation with largest Google search volume	Alternative translation
English	corruption	–
French	corruption	–
Arabic	فساد	–
Spanish	corrupción	corruptela
Portuguese	corrupção	–
(German)	Korruption	–
(Mandarin)	贪污	贿赂 (and others)

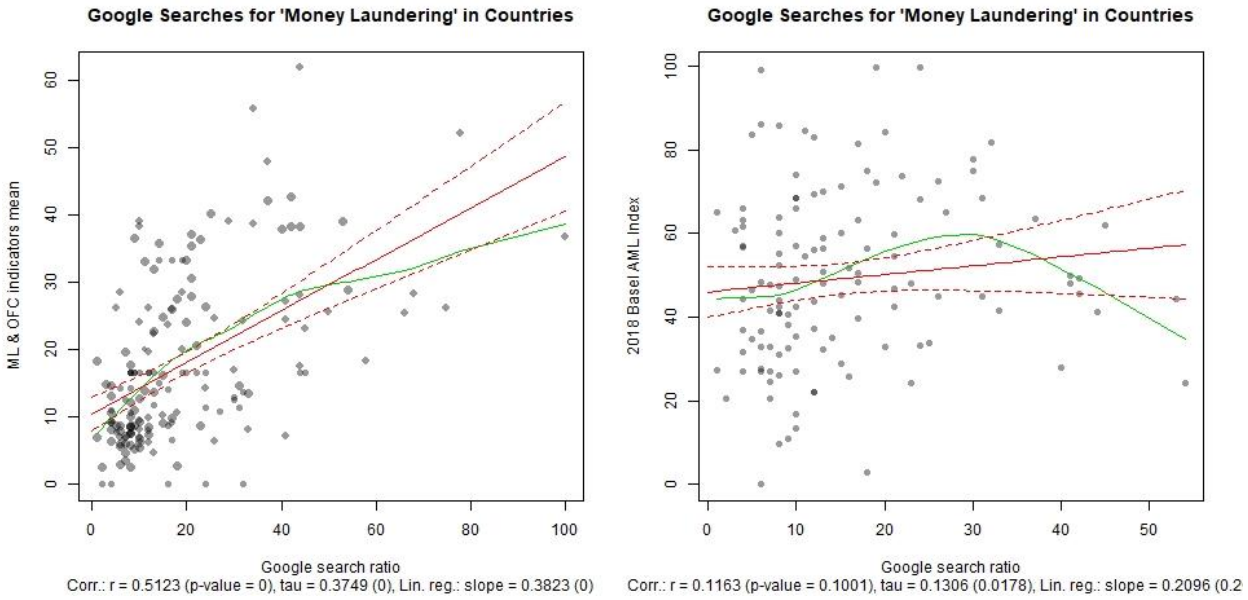
Eventually, following search queries are used as outcome of the described selection process:

"money laundering"+"blanchiment d'argent"+"غسيل الأموال"+"lavado de dinero"+"lavagem de dinheiro"

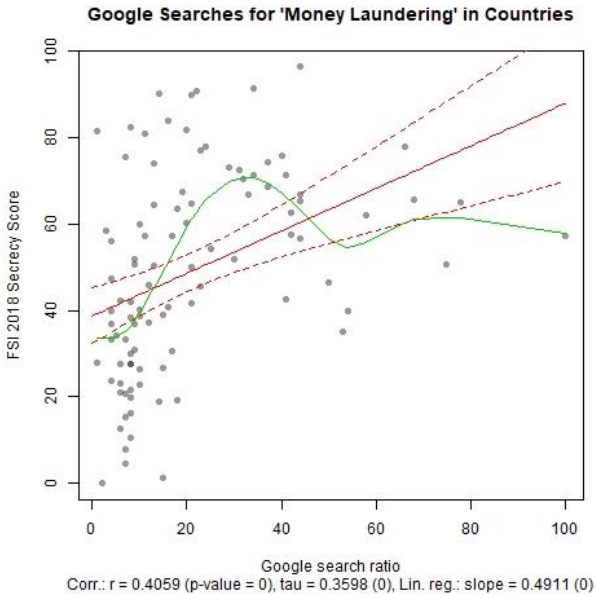
"corruption"+"فساد"+"corrupción"+"corrupção"

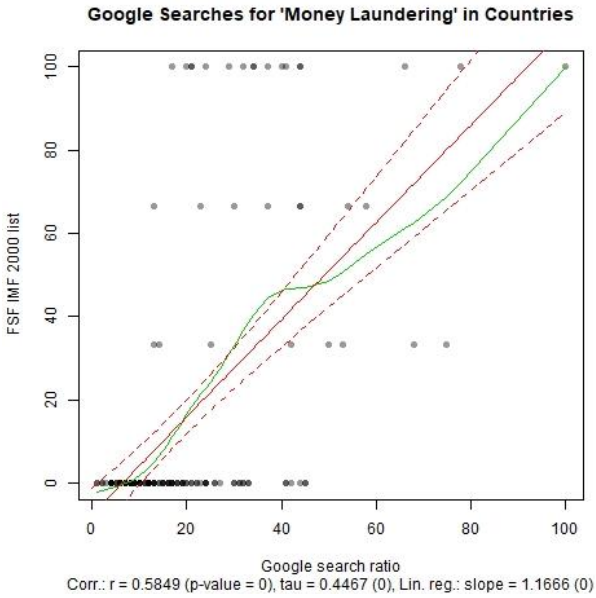
Sources:
https://support.google.com/trends/answer/4359550?hl=en (05.03.2019)
https://support.google.com/trends/answer/4359582?hl=en (05.03.2019)

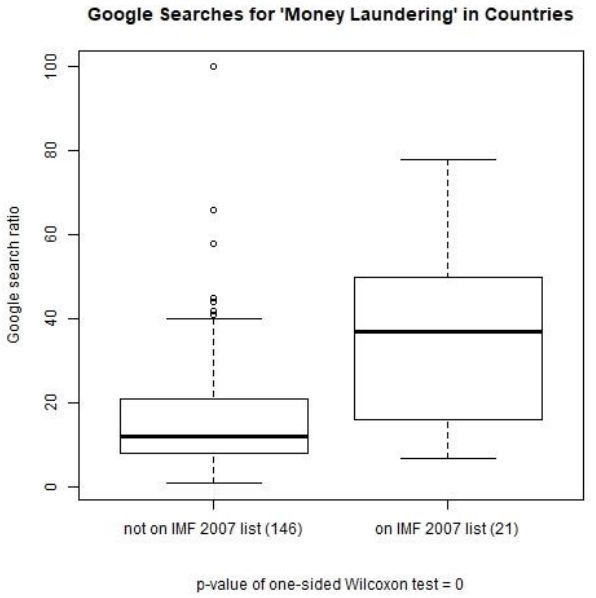
V. Comparison Google search <u>topic</u> *Money Laundering* with individual money laundering indicators

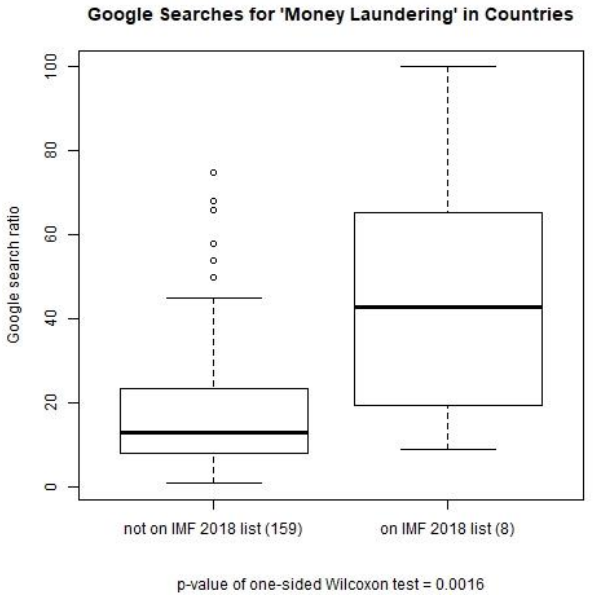


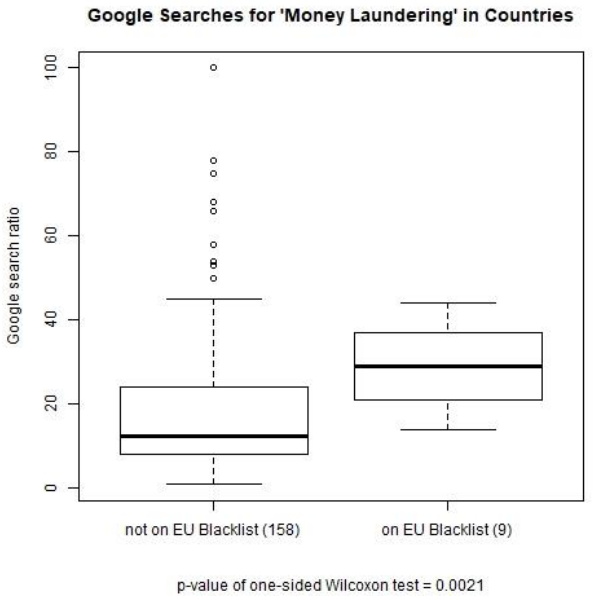
Google Searches Indicate Financial Crime in Countries

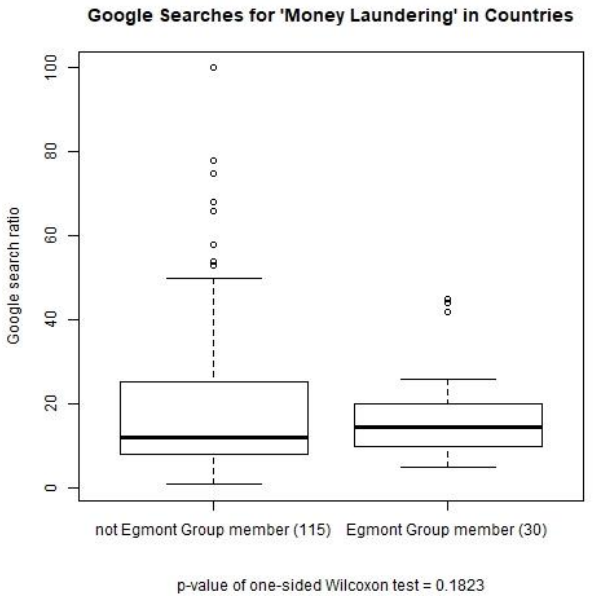


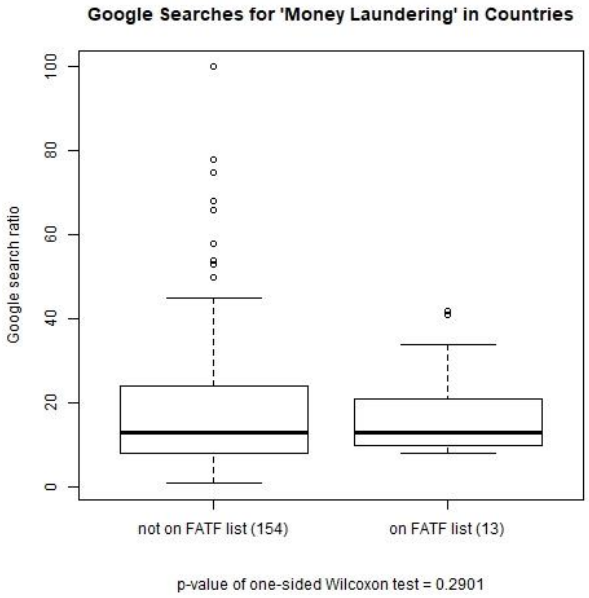








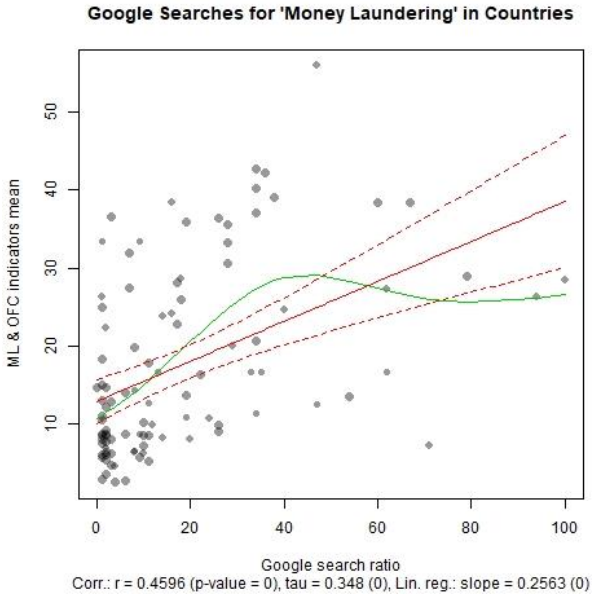


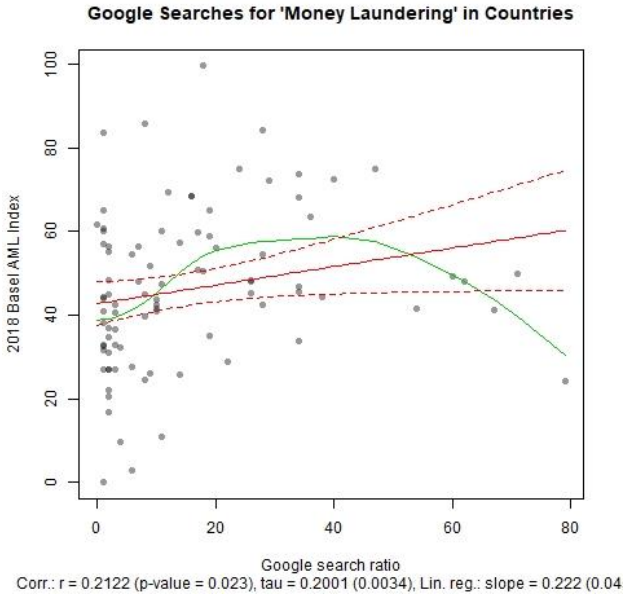


VI. Results for Money Laundering Search Terms

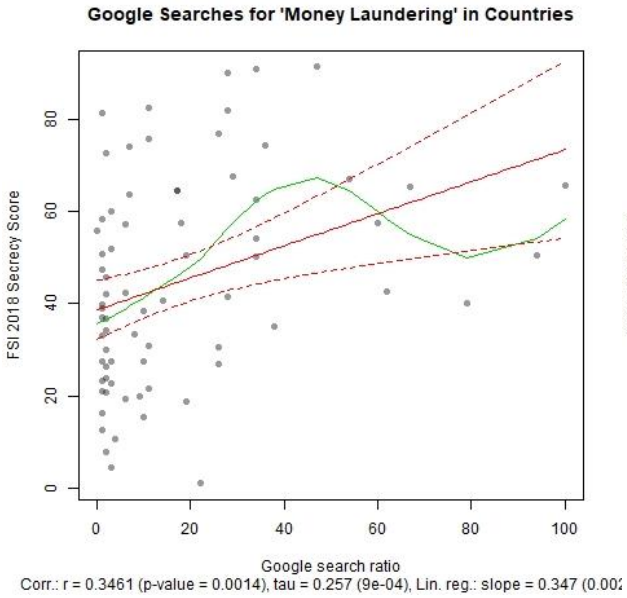
Money Laundering Indicator	Pearson's product-moment correlation		Kendall's rank correlation tau		Wilcoxon test
	r (estimate)	p-value	τ (estimate)	p-value	p-value
Basel AML Index	*0.2122*	*0.023*	*0.2001*	*0.0034*	
FSI Secrecy Score	*0.3461*	*0.0014*	*0.2570*	*0.0009*	
FSF IMF 2000	*0.3829*	*0*	*0.3571*	*0*	
FATF					*0.0784*
Egmont Group					0.1093
EU Blacklist					0.1367
IMF 2007					*0*
IMF 2018					0.0572
mean	*0.4596*	*0*	*0.3480*	*0*	

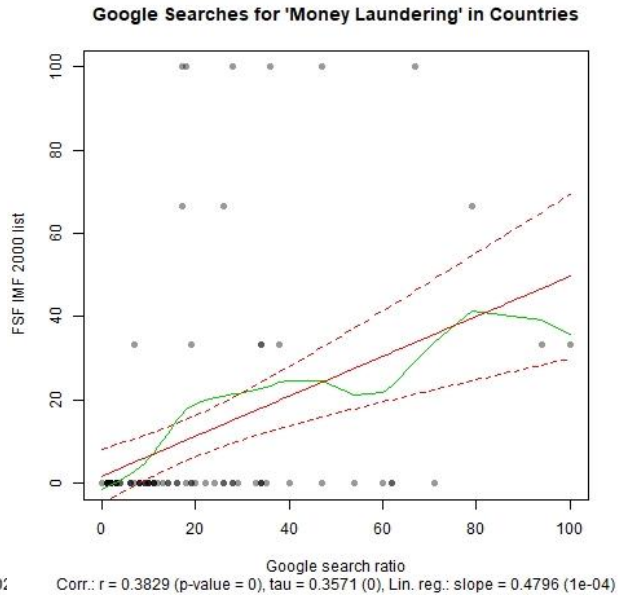
Table 4: Correlation between Google searches on money laundering (search terms) and money laundering indicators.

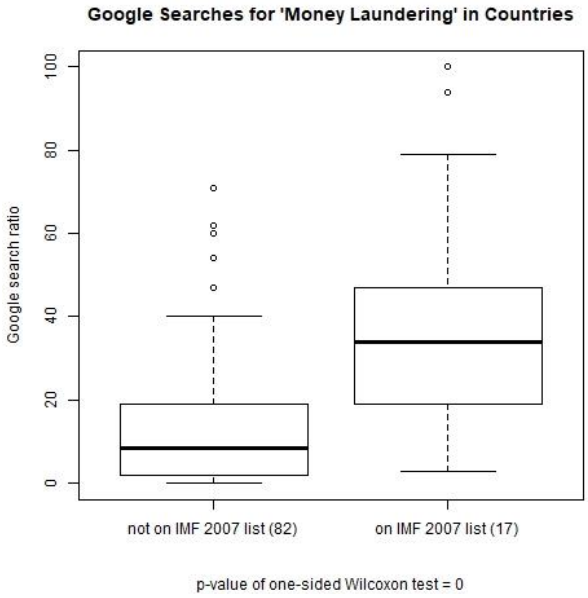


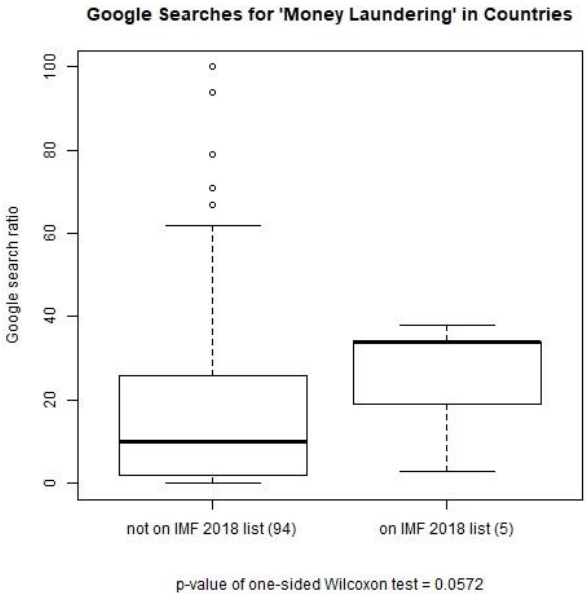


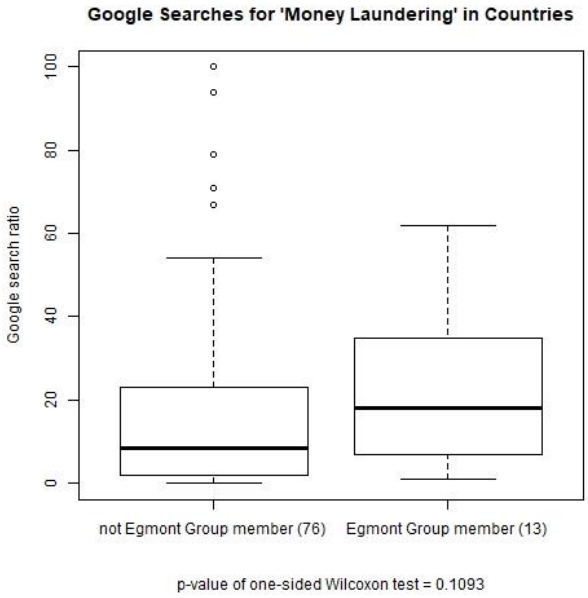
Google Searches Indicate Financial Crime in Countries

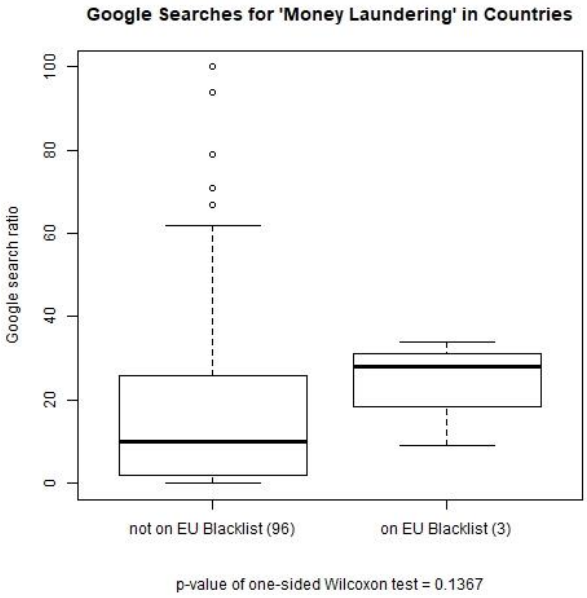


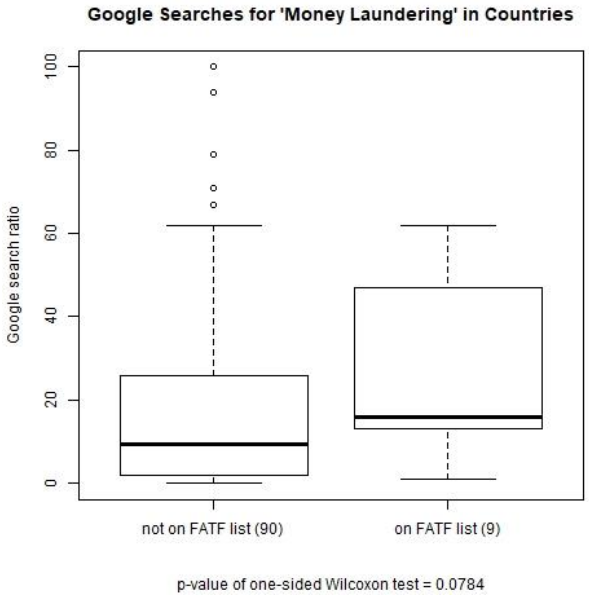




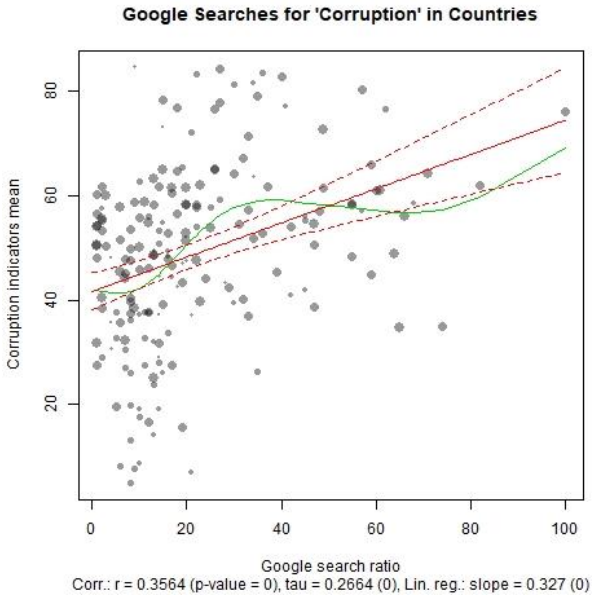


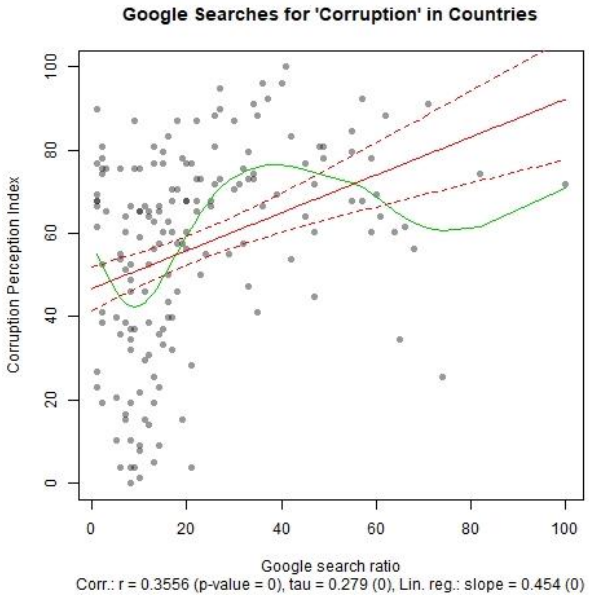


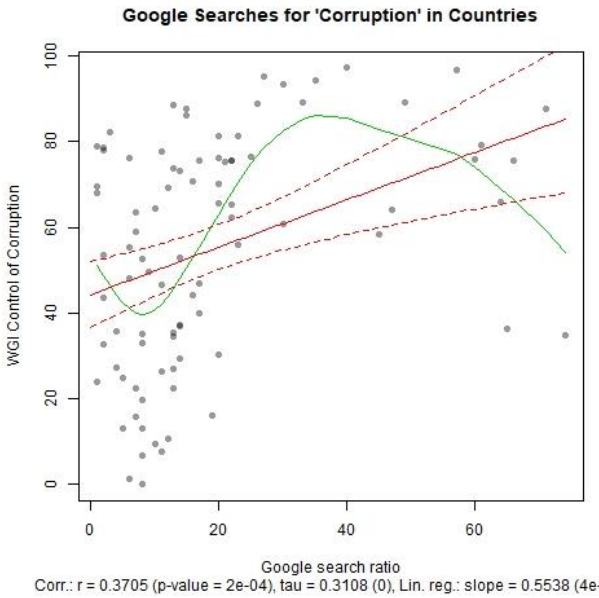


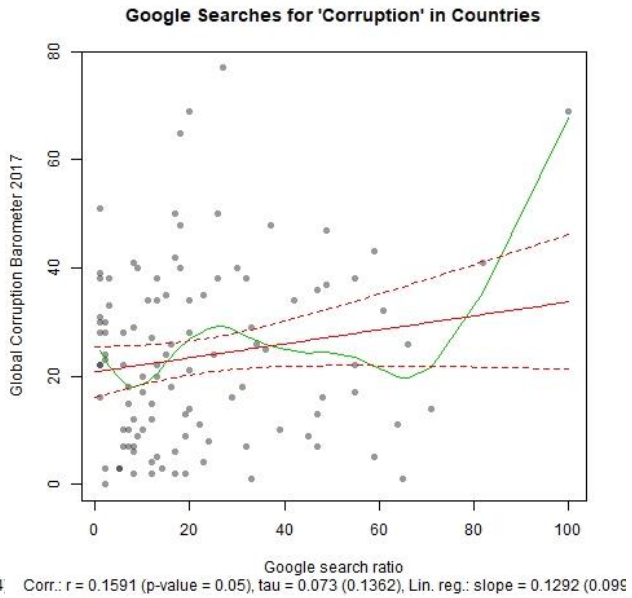


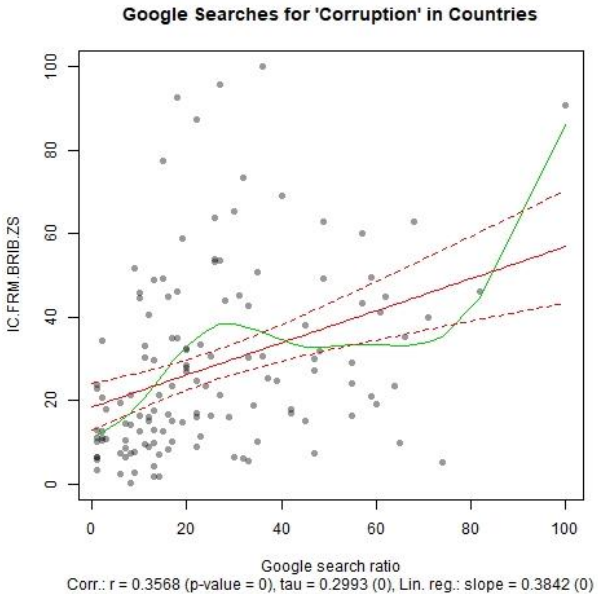
VII. Comparison Google search <u>topic</u> *Corruption* with individual corruption indicators

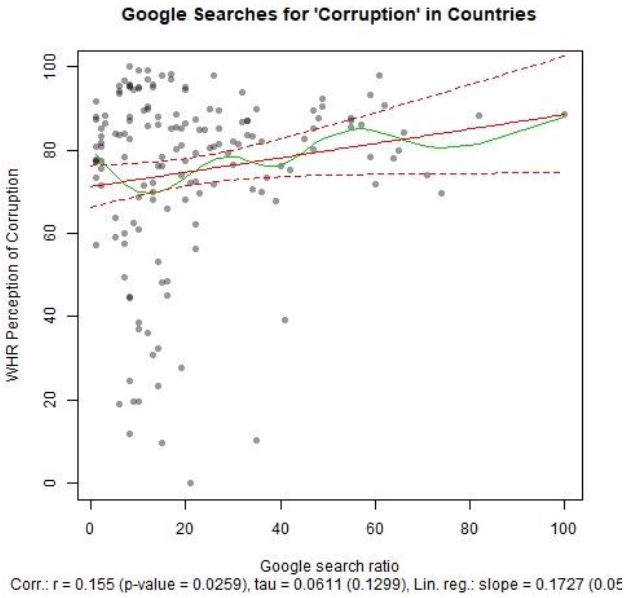


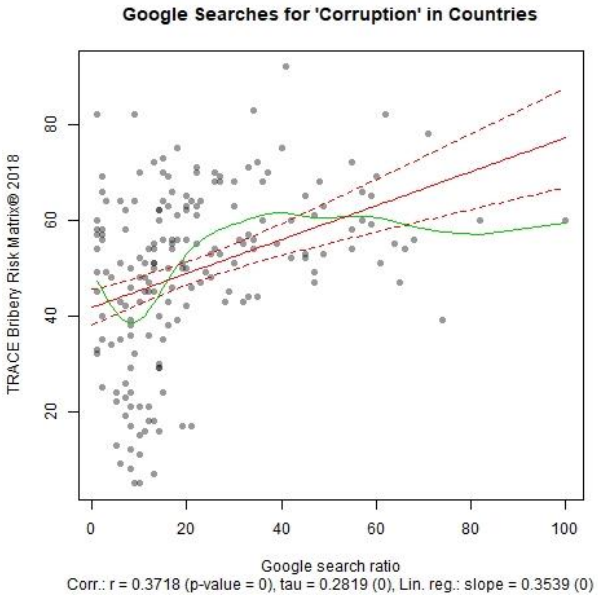










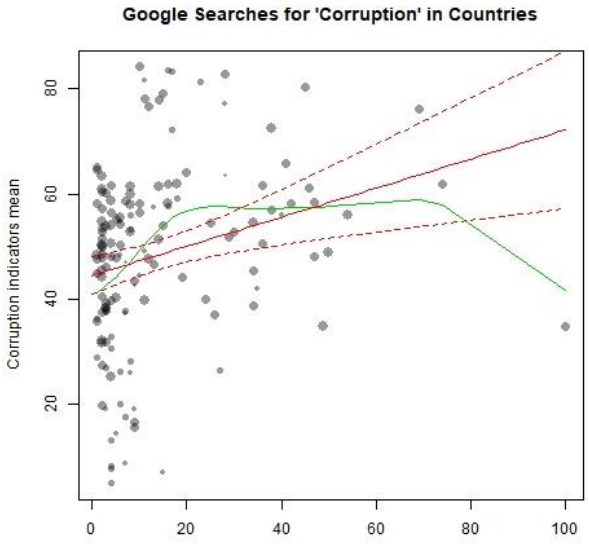


VIII. Results for Corruption <u>Search Terms</u>

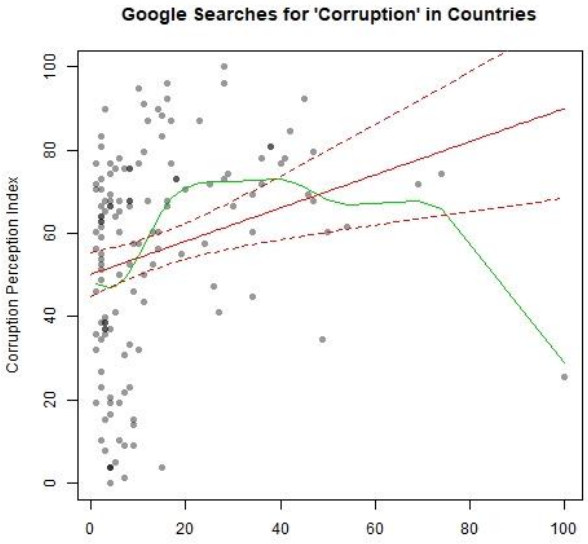
Corruption Indicator	Pearson's product-moment correlation		Kendall's rank correlation tau	
	r (estimate)	p-value	τ (estimate)	p-value
CPI	*0.2614*	*0.0008*	*0.2370*	*0*
WGI Control of Corruption	0.1870	0.0663	*0.1911*	*0.0140*
GCB	0.1515	0.0692	0.0986	0.0838
IC	*0.2095*	*0.0169*	*0.2339*	*0.0003*
WHR	0.0965	0.1309	-0.0638	0.8601
TRACE	*0.3099*	*0.0001*	*0.2483*	*0*
mean	*0.2588*	*0.0008*	*0.2167*	*0.0001*

Table 5: Correlation between Google searches for corruption (<u>search terms</u>) and corruption indicators.

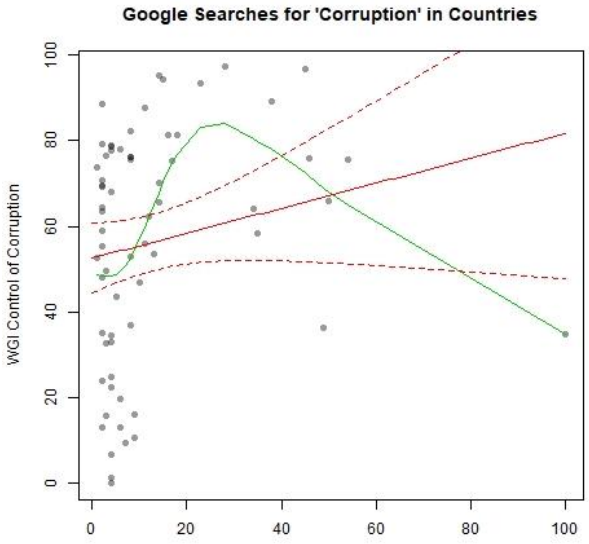
Google Searches Indicate Financial Crime in Countries



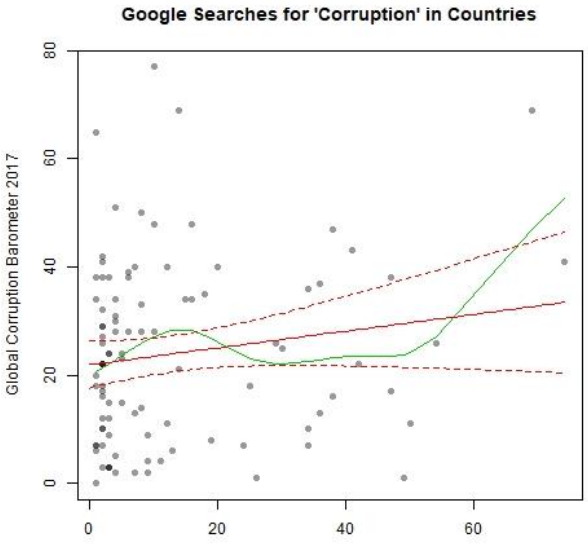
Corr.: r = 0.2588 (p-value = 8e-04), tau = 0.2167 (1e-04), Lin. reg.: slope = 0.2764 (0.00



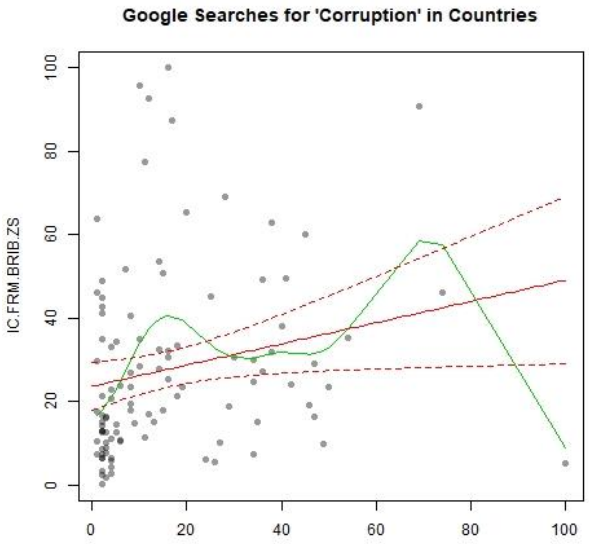
Corr.: r = 0.2614 (p-value = 8e-04), tau = 0.237 (0), Lin. reg.: slope = 0.3979 (0.0017



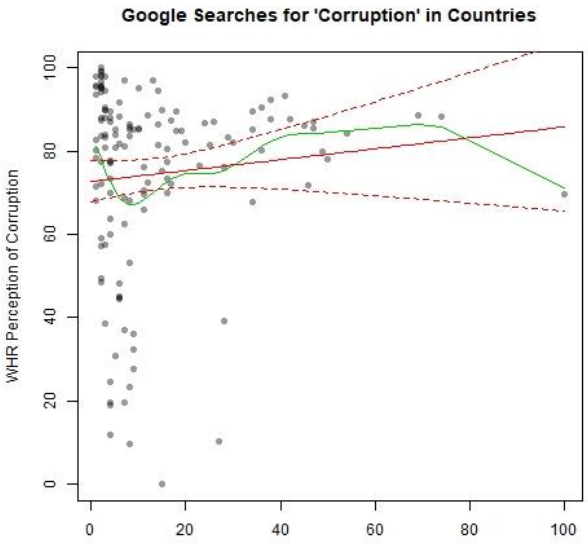
Corr.: r = 0.187 (p-value = 0.0663), tau = 0.1911 (0.014), Lin. reg.: slope = 0.2916 (0.13



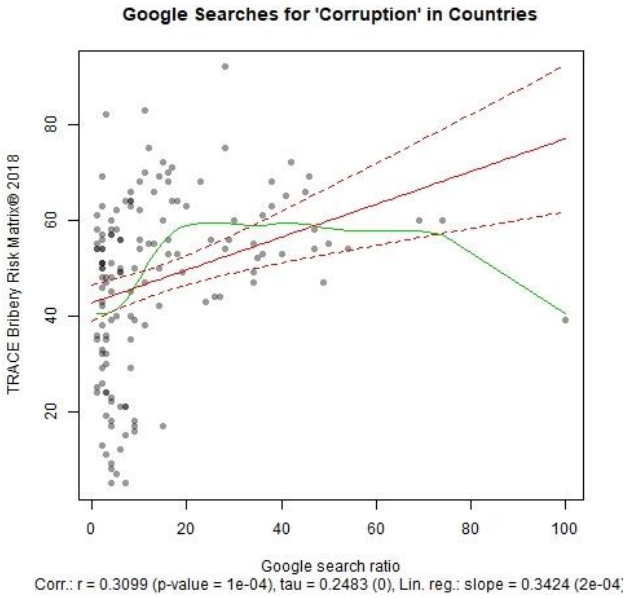
Corr.: r = 0.1515 (p-value = 0.0692), tau = 0.0986 (0.0838), Lin. reg.: slope = 0.1552 (0.1



Corr.: r = 0.2095 (p-value = 0.0169), tau = 0.2339 (3e-04), Lin. reg.: slope = 0.254 (0.03



Corr.: r = 0.0965 (p-value = 0.1309), tau = -0.0638 (0.8601), Lin. reg.: slope = 0.1307 (0.2



IX. References

[1] Seung-Pyo Jun, Hyoung Sun Yoo, San Choi, *Ten years of research change using Google Trends: From the perspective of big data utilizations and applications,* Technological Forecasting and Social Change, Volume 130, 2018, Pages 69-87, ISSN 0040-1625, https://doi.org/10.1016/j.techfore.2017.11.009 (28.04.19)

[2] Futoshi Narita, Rujun Yin: In Search of Information: Use of Google Trends' Data to Narrow Information Gaps for Low-income Developing Countries (2018), IMF Working Paper WP/18/286. https://www.imf.org/~/media/Files/Publications/WP/2018/wp18286.ashx (28.04.19)

[3] J. Ginsberg, M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, L. Brilliant: *Detecting influenza epidemics using search engine query data*, Nature, 457 (2009), pp. 1012-1014. https://doi.org/10.1038/nature07634 (28.04.19)

[4] C. Pelat, C. Turbelin, A. Bar-Hen, A. Flahault, A.-J. Valleron: *More diseases tracked by using Google Trends*, Emerg. Infect. Dis., 15 (2009), pp. 1327-1328. https://dx.doi.org/10.3201/eid1508.090299 (28.04.19)

[5] Preis, T., Moat, H. S., Stanley, H. E., Bishop, S. R.: *Quantifying the Advantage of Looking Forward* (2012), Scientific Reports 2, 350. https://doi.org/10.1038/srep00350 (28.04.19)

[6] Wolfsberg Group: Country Risk Frequently Asked Questions (FAQs) https://www.wolfsberg-principles.com/sites/default/files/wb/Wolfsberg%20FC%20Country%20Risk%20FAQs%20Mar18.pdf (25.04.19)

[7] http://gs.statcounter.com/download/search-country/ (15.02.19)

[8] en.wikipedia.org/wiki/List_of_languages_by_the_number_of_countries_in_which_they_are_recognized_as_an_official_language (17.4.19)

[9] Hamilton, A., Hammer, C.: Can We Measure the Power of the Grabbing Hand? (2018). World Bank Group, Policy Research Working Paper 8299. http://documents.worldbank.org/curated/en/113281515516828746/pdf/WPS8299.pdf (25.04.19)

[10] Rose, A. K., Spiegel, M. M.: Offshore Financial Centers: Parasites or Symbionts? (2006). The Economic Journal: 1310–1335. http://www.nber.org/papers/w12044.pdf (27.04.19)