Investigating High-Risk Firms: A Machine Learning-based Approach to Cross-Border Ownership Data

Antonio Bosisio
Maria Jofre
Transcrime – Università Cattolica Sacro Cuore

Abstract
Corporate ownership secrecy has become a central issue in the global debate as the use of legitimate companies in illicit schemes has increased dramatically in recent times. While several measures have been implemented worldwide to increase the transparency of firms and their owners, empirical evidence and knowledge on the subject remain limited to few small-scale case studies. In addition, there is a lack of tools specifically designed for risk assessment and risk monitoring of firms to be used by public authorities. The present paper, based on the results of the EU-funded project DATACROS, addresses these gaps by (i) proposing and validating novel risk indicators of corporate ownership opacity in a large sample of companies, and (ii) implementing them in a user-friendly platform to be used by public institutions, a tool capable of identifying companies at risk of involvement in corruption and money laundering. Machine learning results confirm the relevance of corporate ownership opacity in the facilitation of financial crime. Firms with (i) more complex structures, (ii) links to secrecy jurisdictions, and (iii) links to opaque corporate vehicles, are, in fact, more prone to engage in illicit activities. This urgently calls for the innovation of risk assessment activities based on the intelligent use of corporate ownership information. As such, the present contribution could be used to support LEAs and other authorities in combating financial crime in the sometimes overwhelming and ever-evolving digital age.

Keywords: financial crime; ownership structure; risk indicators; machine learning; investigative tool

Introduction

Background

Legitimate companies play a crucial role in facilitating corruption schemes and money laundering of illicit proceeds (EFECC, 2020; Europol, 2018; Savona & Riccardi, 2018). Companies are exploited to create a 'screen' that makes it particularly difficult to trace the real identity of the individuals who ultimately control them – the so-called beneficial owners (hereinafter BOs). Recent investigations carried out by European law enforcement agencies (LEAs) and financial intelligence units (FIUs), and recent research highlights important trends in this domain.

First, there is an increased misuse of complex and opaque corporate ownership in illicit schemes that aim to conceal BO information, thus impeding the identification of the individuals who ultimately control a company. According to the World Bank, 70% of corruption cases between 1980 and 2010 involved anonymous shell companies (van der Does de Willebois et al., 2011). Panama Papers (ICIJ, 2016) and Paradise Papers (ICIJ, 2017), among others journalistic investigations, uncovered dense and opaque networks of companies and trusts established to conceal the identity of their beneficial owners, and the criminal origin of their proceeds.

Second, financial crime schemes increasingly exploit cross-border structures: criminals use bank accounts, intermediaries, and firms located in different jurisdictions, including non-cooperative tax havens: in Europe, 1% of limited companies have ownership links with entities coming from blacklisted countries, but in some EU Member States this percentage goes up to 12% (Bosisio et al., 2021).

Third, there is a high volume of cross-links between corruption, organised crime, tax fraud, and money laundering. The outburst of the COVID-19 pandemic, and the introduction of recovery plans by EU Member States, have provided criminal networks with further opportunities to drain public resources through simultaneous use of different financial crime schemes (UNODC, 2020; FATF, 2020).

All these trends exploit weaknesses in the preparedness and capabilities of European law enforcement and judicial authorities to combat financial crimes. Moreover:

- There is a lack of risk assessment tools specifically designed for public authorities: current tools and solutions have been designed primarily for banks, financial institutions, and large corporations (e.g., for anti-money laundering and compliance purposes). There is a dearth of tools specifically designed to support criminal investigations dealing with the monitoring of companies potentially involved in corruption and financial crime. A survey conducted by Transcrime in 2019, involving 37 public authorities from 19 EU countries, including LEAs, FIUs, Anti-Corruption Agencies (ACAs), Competition Authorities (CAs), and Tax Authorities (TAs), revealed that 60% of public authorities do not use software for financial investigations, but 78% would like to have tools for tracing and assessing the risk of firms;
- There is a lack of (i) knowledge and skills for gathering information on companies and related entities/individuals, and (ii) ML-based indicators, models, and tools to identify high-risk companies, also when ownership structures deploy cross-border;
- There is a lack of communication and coordination among stakeholders in the exchange of best practices, investigation, and intelligence practices, and in the implementation of cooperation mechanisms, especially at the transnational level.

In Europe, efforts are being made to facilitate the identification of company owners with the establishment of BO registers, introduced by the fourth (and later fifth) Anti-Money Laundering Directive. However, to have a complete picture of potential risks, it is often not enough to know who controls a company, it is also crucial to understand how control occurs: which shareholding structure is used, what corporate vehicles and jurisdictions are involved, and with what degree of complexity.

Current research

In order to address these gaps, and to increase the knowledge on the issue, the EU-funded projects DATACROS I and II have produced the first analysis of the opacity in the ownership structure of 56 million companies across 29 European countries, and developed the first software for public authorities capable to identify companies at risk of involvement in corruption and money laundering.

The present study, conducted for the purposes of the project, further proposes a two-fold strategy:

1) To define, calculate, and validate relevant ownership risk indicators on corporate secrecy that relate to the three identified facets of opacity, including (i) complex ownership structures, (ii) links to secrecy jurisdictions, and (iii) links to opaque corporate vehicles;

2) To develop a prototype tool for risk assessment purposes based on the proposed secrecy risk indicators.

Methodology

Several risk factors related to features of ownership structures have been identified from the review of the literature, information that could be exploited to better understand and detect financial crimes. Risk factors include: (i) anomalous complexity of ownership structures, (ii) ownership links to high-risk countries, and (iii) ownership links to opaque corporate vehicles. In order to advance extant knowledge on ownership opacity, we defined and assessed ownership risk indicators associated with these three factors.

Data

The datasets used in the present study were retrieved from different sources, including business ownership data, compliance data, and country black and grey lists.

Business ownership data

Information on 56 million companies across 29 European countries[1] was retrieved from Bureau van Dijk's Orbis Europe.[2] In order to guarantee both cross-country and cross-sector comparability, only limited companies with information on the ownership structure were included in the analysis. Consequently, the exploited dataset provided a snapshot of ownership information during the month of June 2019, containing information on 13.4 million companies, and about 20 million BOs.

Sanctions and enforcements

Information on companies and their owners that were either included in a sanction list, or associated with enforcement cases from 9 countries[3] were obtained from LexisNexis WorldCompliance.[4] This included information on companies and business owners reported in: (i) one or more of the global screening and sanction lists issued by the EU, US Office of Foreign Assets Control (OFAC), United Nations (UN), Bank of England, US Federal Bureau of Investigation (FBI), and US Bureau of Industry and Security (BIS), or (ii) associated with enforcement provisions (e.g. arrests, final judgments), and court filings around the world, data collated from various sources including national law enforcement reports, press releases, and other statements from public authorities.[5]

Country blacklists

To operationalise the concept of high-risk jurisdictions, we considered the following black and grey lists:[6]

- *Tax domain*: EU list of non-cooperative jurisdictions for tax purposes, which groups together countries that encourage abusive tax practices, and ultimately erode corporate tax revenues of EU Members States (European Commission, 2019);
- *AML/CTF domain*: FATF lists of non-cooperative jurisdictions (or jurisdictions under increased monitoring) in the global fight against money laundering and terrorist financing (FATF 2019). In particular, two lists were included: (i) Call for action (or so-called 'black list') that identifies countries that are considered by the FATF as non-cooperative in the global fight against money laundering and terrorist financing, who are flagged as 'Non-Cooperative Countries or Territories' (NCCTs), and (ii) Other monitored jurisdictions (or so-called 'grey list') comprising jurisdictions that have strategic AML/CFT deficiencies for which they have developed an action plan together with the FATF (FATF, 2019; 2017).

Risk indicators

For all the companies in the sample, the full ownership structure was reconstructed (Figure 1). For each firm, entities owning more than 10% of the share capital at each ownership level were identified. This process continued until we reached an individual ultimate beneficiary at the top of the chain (i.e. a BO). If it was not possible to identify an individual at the top of a chain, then the top shareholder was referred to as Other Ultimate Beneficiary (OUB). Entities separating a company from its ultimate beneficiaries, either BOs or OUBs, were labelled as intermediate shareholders (INTs).

Each of the proposed risk indicators were measured and operationalised as described below.

Beneficial ownership complexity (BOC)

The first analysed risk factor related to the anomalous complexity of corporate ownership structures. The complexity of an ownership structure was operationalised using the so-called *BO distance*, that is, the number of steps that separate a company from its BO(s). When the *BO distance* is equal to 1, then the company is directly controlled by its BO(s). The greater the BO distance, the higher the level of complexity of the company's ownership structure, hence the more difficult it is to trace its BOs, which in turn represents a greater risk that the company can be used to hide criminal profits and/or individuals (Knobel, 2021).

1 Countries included: EU27 + the United Kingdom + Switzerland.

2 https://www.bvdinfo.com/en-gb/ (last visited: August, 2022)

3 Belgium, Cyprus, France, Italy, Luxembourg, Malta, the Netherlands, Spain, and the United Kingdom.

4 For more information, see https://risk.lexisnexis.com/global/en/products/worldcompliance-data (last visited: August, 2022).

5 For the purposes of our analysis, all categories of crimes and predicate offences covered by LexisNexis were included.

6 For a full list of black and grey listed countries, see ***Annex 1: Black and grey lists considered in the study.***

Figure 1 – Illustration of the different actors of the ownership structure of a company (CO), which includes Beneficial Owners (BOs), Other Ultimate Beneficiaries (OUBs), and intermediate shareholders (INTs).

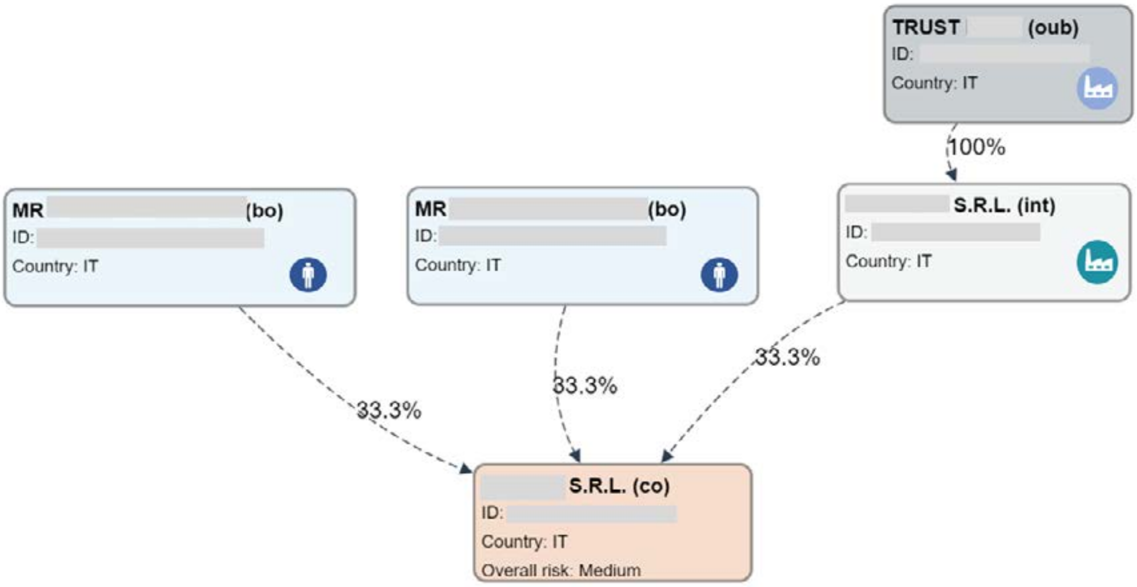
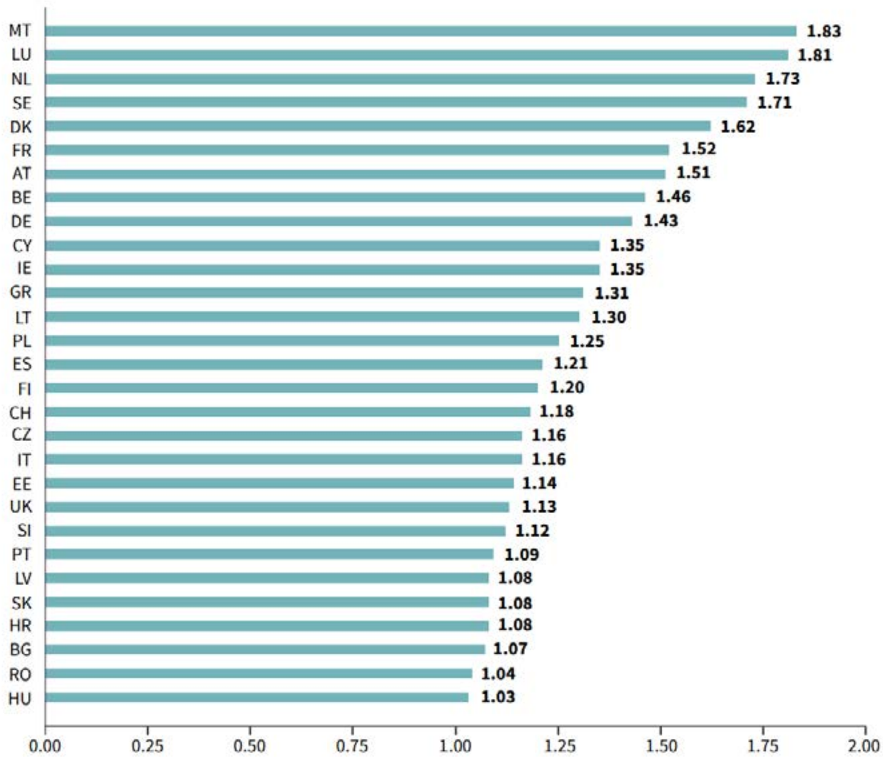


Figure 2 – Average BO distance across European countries (EU27 + UK + CH, 2019)



The average *BO distance* was calculated for all the companies in the sample, and the average observed values were computed at both the territory and sector level. While the average EU value of the BOC indicator was 1.21, significant differences can be observed across countries (Figure 2). Malta was the country that displayed the highest average BO distance among European countries (1.83), followed by Luxembourg (1.81), the Netherlands (1.73), and Sweden (1.71). Conversely, the lowest values were observed in Hungary (1.03), Romania (1.04), and Bulgaria (1.07). Moreover, the analysis conducted at a sector level (NACE rev.2 division) showed that some of the business sectors with the highest density of anomalous complex companies included Water transport (NACE division 50), and Gambling and betting activities (NACE division 92), which aligns with previous research (Savona and Riccardi 2018; 2017), and police investigations (DIA 2019; 2017; 2016).

Beneficial ownership secrecy (BOS)

When a company has ownership links to countries with high levels of secrecy, it is more difficult to trace BOs, hence to carry out financial investigations. Therefore, the greater the number of links to high-risk jurisdictions, the greater the risk that these companies may be misused for criminal purposes (Tax Justice Network, 2015; Tavares, 2013). Consequently, ownership data were matched with black and grey lists of risky jurisdictions issued by EU, and FATF. Then, the number of entities (i.e., BOs, OUBs, INTs) that were linked to risky countries for all the ownership structures under study were estimated.

Results showed that the average percentage of companies with ownership connections to black/grey listed jurisdictions across the EU is 0.91%. Furthermore, Luxembourg (8.7%), and Cyprus (8.5%) were by far the countries with the highest values (Figure 3), while the lowest estimates were observed in Portugal (0.1%), Estonia (0.2%), and Slovenia (0.2%). Interestingly, it can be seen that in some countries, such as Belgium, Switzerland, and the United Kingdom, a relevant portion of the links to blacklisted countries were to BOs (i.e., individuals), whilst in others, such as Cyprus, Luxembourg, and the Netherlands, the largest major proportion of these links were not related to individuals, but rather to other firms that were intermediate companies (i.e. firms somewhere in the ownership chain between the company at issue and its BOs), or other ultimate beneficiaries (i.e. firms and corporate vehicles that are at the top of an ownership chain, and do not allow for the identification of the BOs).

Beneficial ownership unavailability (BOU)

In some cases, the identification of the BO(s) of a company is not possible. This may be due to a highly fragmented share capital structure where no one individual owns more than 10% of the shares, or because certain specific corporate vehicles are used deliberately to conceal the identity of individuals at the top of the ownership chain. While the first case of fragmented structures is perfectly legal, and in some contexts even common, the latter option represents a risk factor since the more difficult it is to correctly identify the BOs, the higher the risk that the company can be used to conceal illicit activities. As such, we defined and calculated the BOU indicator for each of the companies as the number of ultimate owners, if any, that are an opaque corporate entity, including trusts, fiduciaries, foundations, and investment funds, which, by statute, do not allow for the identification of the BO(s).

Across the EU, on average, 1.45% of companies were controlled by a trust, a fiduciary, or a fund. As illustrated in Figure 4, the analysis outlined high values in the Netherlands, where 25.6% of the limited companies in our sample were in fact controlled by an opaque corporate vehicle. This is most likely connected to the extended domestic use of Dutch foundations (so-called *stichting*), which are legal arrangements exploited for a range of legitimate purposes: in the Netherlands are commonly used to control for-profit limited or unlimited firms. However, given their specific nature, it is not very meaningful to talk about ‘owners’ of a *stichting*, and for this purpose they may be misused to hide the identity of the ultimate beneficiaries (OECD, 2019).

Processing of risk indicators

A final processing of the proposed risk indicators involved the transformation from continuous values to risk scores. To this end, we separated the sample into groups of peer companies (so-called *peer groups*), that is, groups of companies active in the same business sector and with a comparable dimension, and further classified companies into five non-overlapping classes using a K-means hierarchical clustering algorithm. This resulted in each company in the sample being assigned a BOC, BOS, and BOU risk score ranging from 1 to 5: the greater this value, the higher the level of risk.

Correlation among risk indicators

As depicted in Table 1, all three ownership indicators showed a positive correlation with each other at the country, regional (nuts2), and sectoral (NACE rev.2 division) levels. The strongest correlation coefficients were observed at country level (a.), while smaller but still significant correlations were observed at regional level (b.). On the contrary, little to no dynamics were observed at sector level. These results suggest that each of the risk indicators captures different facets of corporate ownership features, and that the concentration of anomalous companies seems to be driven by country level-dynamics, such as national legislations and regulations, rather than by industry-driven factors.

Table 1 – Pearson correlation among ownership indicators at a. country level, b. sub-country level (NUTS2), and c. sector level (NACE rev.2 division)

	BOC	**BOS**	**BOU**
BOC	1		
BOS	a. 0.52*** b. 0.46*** c. 0.22**	1	
BOU	a. 0.78*** b. 0.58*** c. 0.07	a. 0.36* b. 0.23*** c. 0.10	1

Figure 3 – Percentage of companies with ownership links to black/grey listed jurisdictions (EU27 + UK + CH, 2019)

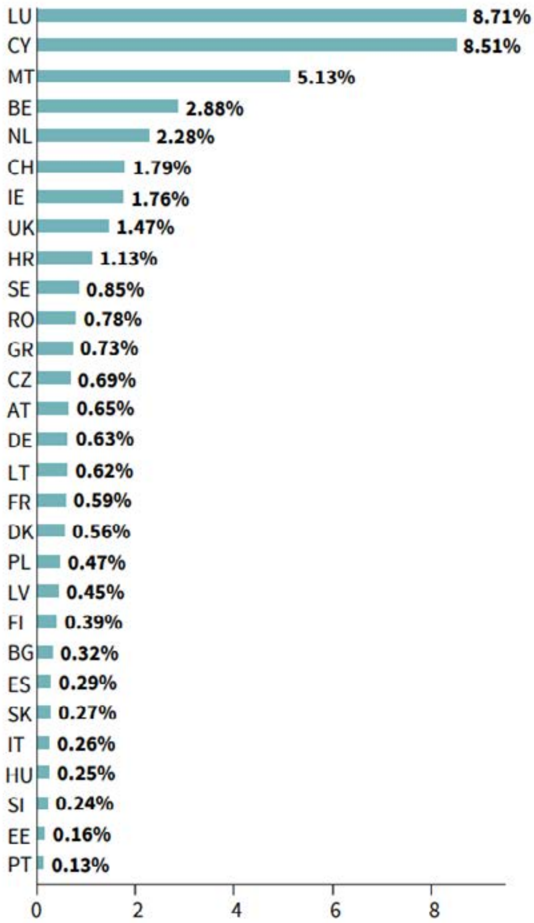
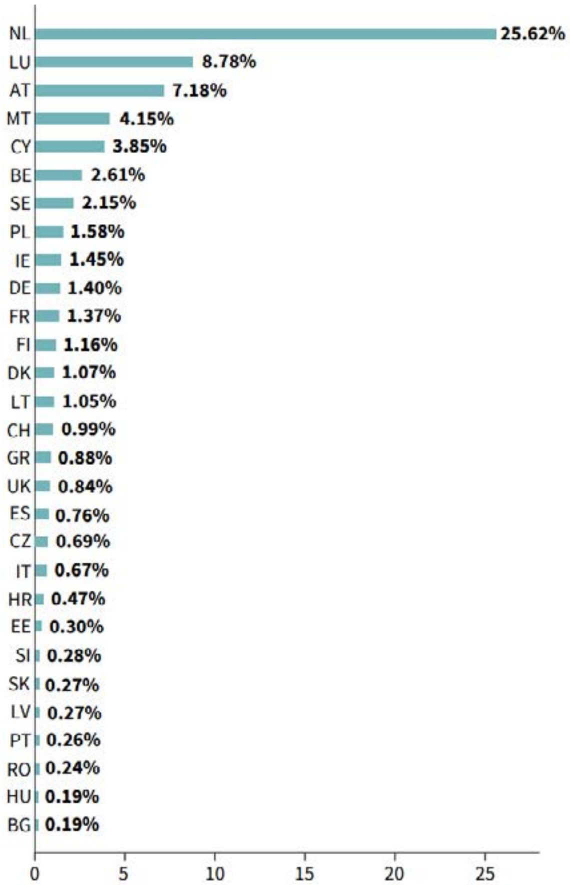


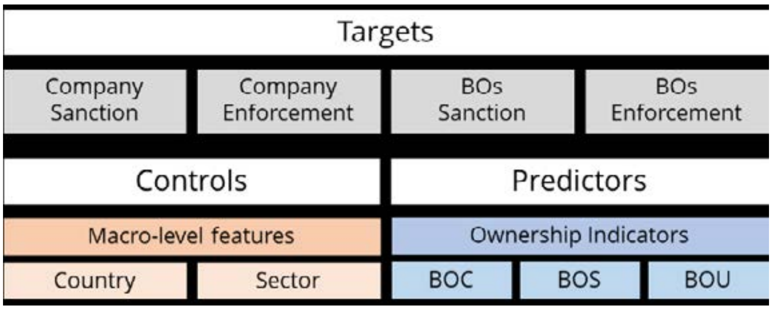
Figure 4 – Percentage of companies with ownership links to opaque corporate vehicles (EU27 + UK + CH, 2019)



Validation of indicators

The proposed risk indicators were then validated by training and testing various machine learning models, thus establishing their usefulness to identify companies that are potentially involved in illicit activities. For the purposes of validation, a sample of around 3 million limited companies registered in the nine European countries from where enforcement and sanction data was used.[7] In particular, we considered (i) as target variables, sanctions and enforcement flags from LexisNexis WorldCompliance, (ii) as predictors, the proposed ownership risk indicators (i.e., BOC, BOS, BOU), and (iii) as controls, a set of country and sector-level binary variables (Figure 5).

Figure 5 – Variables used for modelling: 4 target variable (sanctions on companies, enforcements on companies, sanctions on BOs, enforcements on BOs), three predictors (BOC, BOS, BOU), and two controls (country, economic sector).



Several machine learning models have been implemented, both for the detection of sanctions and enforcement cases, and for the assessment of the predictive performance of the ownership risk indicators. Machine learning models included logistic regression, decision trees, bagged trees, and random forests. All methods have been fitted using a training set (80% of the sample), and further validated on a test set (20%), which ultimately ensured a non-biased estimation of the predictive ability of both the models, and the risk indicators. To manage the imbalance of the target variables, we employed a simple but effective sampling strategy on the training set based on the under-sampling of the majority class (i.e. non-sanctioned/non-enforced observations) that we randomly matched to the number of observations in the minority class (i.e. sanctioned/enforced observations). A robustness analysis based on logistic regression was also performed to assess the stability of the results when cases from a certain country or business sector are excluded.

Satisfactory performance was achieved by all the considered machine learning methods, particularly regarding sanction offences.[8] In the case of logistic regression (Table 2), the algorithm correctly predicted 83.3% of sanctions on companies, and 88% of sanctions on owners. The prediction of companies and owners not subject to sanctions or prior enforcement was also good. The lowest performance occurred when predicting owners in the UK, who have either been subject to or not subject to enforcement, which is suggestive of a more complex country-specific phenomenon.

7 We separated UK from the sample since the number of observations (both number of companies and sanctions/enforcements) compared to the rest of the countries was extremely large, hence eroding the performance of models. This asymmetry can be explained by the higher coverage of LexisNexis in the UK.

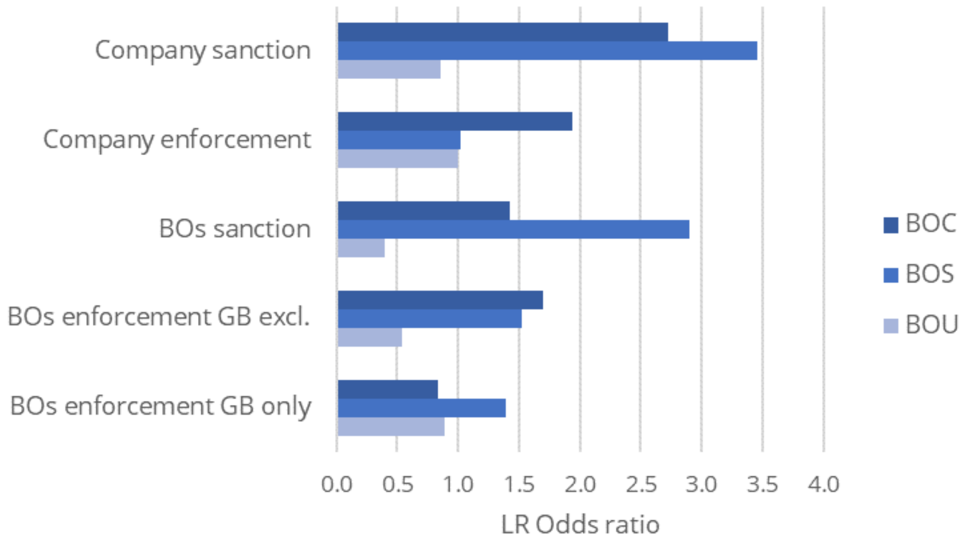
8 More details of machine learning results in Annex 2: Prediction accuracy of different models for the different target variables.

Table 2 – Overall predictive power (true positive and true negative rates) of risk indicators

Logistic regression (test set)	True positive rate	True negative rate
Company sanction	0.833	0.872
Company enforcement	0.679	0.729
BO sanction	0.879	0.851
BO enforcement excl. UK	0.615	0.564
BO enforcement UK	0.548	0.522

Regarding the predictive ability of the indicators, it is observed that BOS was notably important for detecting most offences, particularly with respect to sanction cases (Figure 6). Regarding BOC, there is also evidence of its ability to predict sanctions and enforcement on companies. The BOU indicator appeared to be less relevant in terms of predictive power, but still useful when used collectively.

Figure 6 – Logistic regression odd-ratios of risk indicators by target variable



While the results were stable across the whole sample, some country and sector-specific patterns were observed. For instance, in Italy, Cyprus, and Spain, ownership complexity (BOC) seemed to present a strong connection with illicit behaviour of companies. Ownership links to high-risk jurisdictions (BOS), and ownership links to opaque corporate vehicles (BOU) were more relevant in Malta and the Netherlands. At the sector level, we observed that anomalous ownership complexity (BOC), and ownership links to high-risk jurisdictions (BOS) were major determinants of enforcement and sanction offences in the Financial and insurance sector, while ownership links to opaque corporate vehicles (BOU) was an important factor in the Wholesale and retail trade, as well as Transporting and storage sector.[9]

To conclude, the proposed risk indicators have demonstrated a strong predictive power, confirming that firms with: (i) anomalous complexity of ownership, (ii) ownership links to high-risk jurisdictions, and (iii) ownership links to opaque corporate vehicles, are more prone to engage in illicit activities. Interesting country and sector-specific patterns were observed, evidencing a dynamic and transnational phenomenon, which needs to be tackled by means of innovative technologies, such as the DATACROS tool.

9 For more details see (Author. 2020).

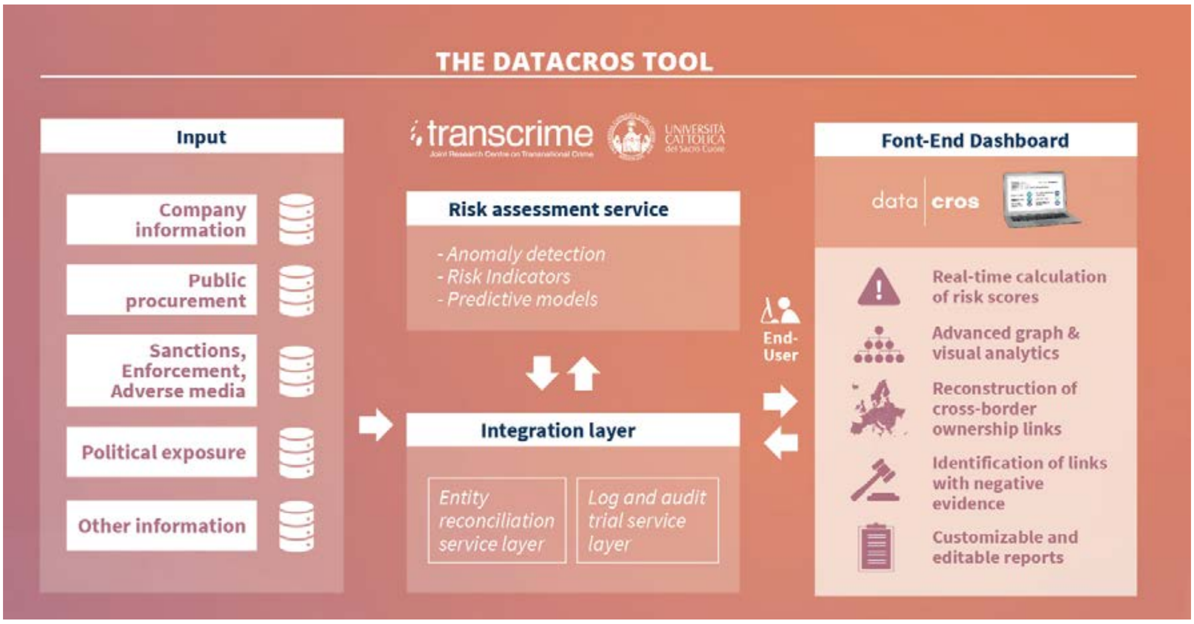
The DATACROS Tool

DATACROS is a research project co-funded by European Union Internal Security Fund – Police, and coordinated by the research centre Transcrime – Università Cattolica del Sacro Cuore, aimed at developing a tool to detect anomalies in firms' ownership structure that can flag high risks of money laundering, collusion, and corruption in the European single market. The first phase of the project (DATACROS I) was conducted between 2019 and 2021 with the participation of the French anti-corruption Authority (Agence Française Anticorruption), the Spanish Police (Cuerpo Nacional de la Policia), and investigative journalists from the IRPI consortium. A second phase of the project (DATACROS II) has started in February 2022 that will last for two years. It will aim at enhancing the Datacros prototype tool, and to test it in operational scenarios with a wide range of end-users, including LEAs, AROs, ACAs, CAs, and investigative journalists. The project consortium, led by Transcrime, is composed by 18 institutions located in 7 different EU countries (Italy, Romania, Spain, France, Belgium, Lithuania, and Czech Republic), including also international organisations and global networks, such as Europol and the Network of Corruption Prevention Authorities (NCPA). For more information, visit: https://www.transcrime.it/datacros/.

DATACROS is only one of several projects of the TOM – The Ownership Monitor research group, a joint initiative recently launched by Transcrime together with its spin-off Crime&tech, to study the opacity of corporate structures in Europe (and beyond).

Project DATACROS I has developed a prototype tool for risk assessment of legitimate companies, able to detect anomalies in firms' ownership structure that can flag high risks of collusion, corruption, and money laundering. This prototype tool is a real-time analytical platform that can be used to investigate anomalies in EU firms' ownership structures, and to conduct risk assessments. The tool complements traditional approaches (e.g. sanctions list checks) with innovative machine learning algorithms, such as the ones presented in the previous sections of this study. In particular, the tool allows to:

- Trace and reconstruct cross-border links among companies, individuals, and related entities (i.e., BOs, shareholders, directors);
- Calculate risk indicators at firm-level in real time, in order to orient, target, and prioritise investigations;
- Detect cartels and clusters of firms that may signal collusive behaviour;
- Identify links with firms and individuals targeted by sanctions and enforcement;
- Visualise graph, maps, and dynamic analytics components to simplify screening activities.



During the second phase of the project (DATACROS II, 2022-2024), the tool will be empowered, fully deployed, and validated by a wider set of public authorities (i.e., LEAs, AROs, ACAs, CAs, and investigative journalists) in different operational scenarios.

In particular, the tool will integrate:

- A wider set of risk indicators, suggested by the Project Consortium, such as financial anomalies, anomalous geographic concentrations, anomalies in turnover of owners and directors, and links to Free Trade Zones;
- New data sources (e.g., company financials, procurement data, sanctions and enforcement data, PEPs) with global coverage (200 countries, 300+ million firms), allowing to trace complex networks, also beyond EU borders;
- New risk assessment functionalities, and machine learning-based entity resolution algorithms;
- Enhanced IT security and personal data protection architecture, to ensure its compliance with governing laws at EU and national level (e.g. Directive 680/2016 and GDPR).

Conclusions

Due to the increased use of legitimate companies in illicit schemes, corporate ownership secrecy has become a central issue in the global political and economic debate. While several measures have been implemented worldwide to increase the transparency of firms and their owners, empirical evidence and knowledge on the subject remains limited to few case studies: there is a complete absence of large-scale analyses. Moreover, there is a lack of tools that are specifically designed for risk assessment and risk monitoring of firms to be used by public authorities (e.g., LEAs, FIUs, CAs, ACAs, TAs).

Schemes are getting more complex (e.g., cross-border, use of opaque vehicles, complex ownership schemes), but information is getting richer. Therefore, advance methodologies are required to prepare LEAs and other authorities as to adequately combat financial crime in the digital age. It is fundamental to develop knowledge and skills to support: (i) gathering of information on companies and related entities/individuals, (ii) developing of ML-based indicators and models to identify high-risk companies, and (iii) implementation of customised tools for investigation and risk assessment of companies and owners. In fact, current tools and solutions available on the market are designed primarily for financial institutions (e.g. for anti-money laundering and compliance purposes), revealing a lack of tools specifically designed for public authorities.

In response to this, we propose and validate an innovative analytical approach for measuring the opacity of corporate ownership through a set of secrecy risk indicators. The proposed risk indicators have demonstrated a strong predictive power, confirming the relevance of corporate ownership opacity as a key element to fight financial crime. The analysis conducted indicates that even strong and stable economies within the EU are vulnerable in this regard. Firms with (i) anomalous complexity of ownership, (ii) ownership links to high-risk jurisdictions, and (ii) ownership links to opaque corporate vehicles are, in fact, more prone to engage in illicit activities.

The present study also presents the DATACROS tool, a prototype software that allows to calculate in real time the ownership risk indicators discussed in this paper, integrating them in an analytical platform designed to support financial crime investigations by public authorities. In the first phase of the project

(2019-2021), the tool has been tested by different end users, including the French Anticorruption Agency, the Spanish Police, and the investigative journalists from IRPI, who have reported a high level of satisfaction with the tested tool. A second phase of the project (DATACROS II) has started in February 2022, and will last for two years. It will aim to enhance the DATACROS prototype tool, and to test it in operational scenarios with a wide range of end-users: LEAs, AROs, ACAs, CAs, and investigative journalists. The project consortium, led by Transcrime, is composed by 18 institutions located in 7 different EU countries (Italy, Romania, Spain, France, Belgium, Lithuania, Czech Republic), including also international organisations and global networks, such as Europol, and the Network of Corruption Prevention Authorities (NCPA).

The findings of the present research lead us to suggest various recommendations. First, it is required to improve the assessment and mapping of high-risk areas and sectors of activity, and how this impacts the misused of legitimate structures by organised crime, and other criminal actors. Improving the monitoring exercise could only enhance understanding of how risks evolve and change, overall and across territories and industries.

Second, there is a growing need for data analytics solutions and risk indicators to increase the effectiveness of monitoring and supervision of ownership opacity.

The last recommendation relates to the improvement of information exchange and cooperation among public authorities. As the latest SOCTA report highlighted, current criminal schemes entail crosslinks among corruption, money laundering, organised crime, and tax fraud (Europol 2021). This calls for the EU to support activities that promote communication, coordination, and cooperation among the wide variety of stakeholders active in the fight of corruption, money laundering, and other financial crimes, including LEAs, ACAs, CAs, FIUs, TAs, investigative journalists, and civil society NGOs.

Acknowledgements

The authors would like to acknowledge that the present study results from the research activity of DATACROS, project funded by the European Union Internal Security Fund – Police (ISFP-2017-AG-CORRUPT-823792).

Annexes

Annex 1: Black and grey lists considered in the study

Updated as of October/November 2019

List	Countries included
EU black list of non-cooperative jurisdictions for tax purposes (08/11/2019)	American Samoa, Fiji, Guam, Oman, Trinidad and Tobago, United States Virgin Islands, Vanuatu, Samoa
EU grey list of non-cooperative jurisdictions for tax purposes (08/11/2019)	Anguilla, Antigua and Barbuda, Armenia, Australia, Bahamas, Barbados, Bermuda, Bosnia and Herzegovina, Botswana, Belize, British Virgin Islands, Cape Verde, Cayman Islands, Cook Islands, Curacao, Jordan, Maldives, Marshall Islands, Mongolia, Montenegro, Morocco, Namibia, Nauru, Niue, Palau, Saint Kitts and Nevis, Saint Lucia, Seychelles, Swaziland, Thailand, Turkey, Vietnam
FATF AML black list (October 2019 statement) – Call for action	Iran, Democratic People's Republic of Korea
FATF AML grey list (October 2019 statement) – Other monitored jurisdictions	Bahamas, Bouvet Island, Cambodia, Ghana, Iceland, Mongolia, Palau, Papua New Guinea, Tajikistan, Tunisia, Yemen, Zimbabwe

Annex 2: Prediction accuracy of different models for the different target variables

Accuracy metrics include true positive rate (TPR), true negative rate (TNR), overall accuracy, and area under the curve (AUC)

	TPR	TNR	Accuracy	AUC
	Logistic Regression (LR)			
Company sanction	0.833	0.872	0.853	0.931
Company enforcement	0.679	0.729	0.704	0.785
BOs sanction	0.879	0.851	0.865	0.896
BOs enforcement UK excl.	0.615	0.564	0.589	0.634
BOs enforcement UK only	0.548	0.522	0.535	0.550
	Decision Tree (DT)			
Company sanction	0.876	0.846	0.861	0.919
Company enforcement	0.769	0.634	0.701	0.731
BOs sanction	0.869	0.856	0.863	0.879
BOs enforcement UK excl.	0.520	0.675	0.598	0.639
BOs enforcement UK only	0.874	0.164	0.524	0.533
	Bagged Trees (BT)			
Company sanction	0.910	0.778	0.844	0.918
Company enforcement	0.763	0.634	0.698	0.759
BOs sanction	0.856	0.841	0.849	0.890
BOs enforcement UK excl.	0.515	0.670	0.592	0.640
BOs enforcement UK only	0.874	0.164	0.524	0.533
	Random Forest (RF)			
Company sanction	0.752	0.868	0.810	0.922
Company enforcement	0.729	0.662	0.696	0.766
BOs sanction	0.851	0.859	0.855	0.885
BOs enforcement UK excl.	0.578	0.610	0.594	0.649
BOs enforcement UK only	0.550	0.523	0.537	0.554

References

- Bosisio, A., Nicolazzo, G. & Riccardi, M. (2021) *The changes in ownership of Italian companies during the Covid-19 emergency*. Milano: Transcrime – Università Cattolica del Sacro Cuore (Transcrime Research in Brief, 5). Available at: https://www.transcrime.it/en/publications/the-changes-in-ownership-of-italian-companies-during-the-covid-19-emergency/ (Accessed: 22 June 2021)
- DIA (2016) *Relazione semestrale sull'attività svolta e sui risultati conseguiti dalla Direzione investigativa antimafia – secondo semestre 2016*. Ministero dell'Interno.
- DIA (2017) *Relazione semestrale sull'attività svolta e sui risultati conseguiti dalla Direzione investigativa antimafia – secondo semestre 2017*. Ministero dell'Interno.
- DIA (2019) *Relazione semestrale sull'attività svolta e sui risultati conseguiti dalla Direzione investigativa antimafia – secondo semestre 2019*. Ministero dell'Interno.
- EFECC (2020) *Enterprising criminals – Europe's fight against the global networks of financial and economic crime*. Available at: https://www.europol.europa.eu/publications-documents/enterprising-criminals-%E2%80%93-europe%E2%80%99s-fight-against-global-networks-of-financial-and-economic-crime
- European Commission (2019) Evolution of the EU list of tax havens. Available at: https://ec.europa.eu/taxation_customs/sites/taxation/files/eu_list_update_08_11_2019_en.pdf.
- Europol (2018) EU-wide VAT fraud organised crime group busted, Europol. Available at: https://www.europol.europa.eu/newsroom/news/eu-wide-vat-fraud-organised-crime-group-busted (Accessed: 5 January 2021)

- Europol (2021) EU Serious and Organised Crime Threat Assessment 2021. The Hague: EUROPOL.
  Available at: https://www.europol.europa.eu/activities-services/main-reports/european-union-serious-and-organised-crime-threat-assessment
- FATF (2017) FATF Report to the G20 on Beneficial Ownership. Paris: Financial Action Task Force.
  Available at: http://www.fatf-gafi.org/publications/mutualevaluations/documents/report-g20-beneficial-ownership-2016.html (Accessed: 25 January 2017)
- FATF (2019) Improving Global AML/CFT Compliance: On-going Process – 18 October 2019.
  Available at: http://www.fatf-gafi.org/publications/high-risk-and-other-monitored-jurisdictions/documents/fatf-compliance-october-2019.html (Accessed: 5 January 2021)
- FATF (2020) COVID-19-related Money Laundering and Terrorist Financing Risks and Policy Responses.
  Available at: https://www.fatf-gafi.org/media/fatf/documents/COVID-19-AML-CFT.pdf
- ICIJ (2016) About the Panama Papers Investigation, ICIJ.
  Available at: https://www.icij.org/panama-papers-about-the-investigation/ (Accessed: 5 January 2021)
- ICIJ (2017) Paradise Papers Exposes Donald Trump-Russia links and Piggy Banks of the Wealthiest 1 Percent, ICIJ.
  Available at: https://www.icij.org/investigations/paradise-papers/paradise-papers-exposes-donald-trump-russia-links-and-piggy-banks-of-the-wealthiest-1-percent/ (Accessed: 5 January 2021)
- Jofre, M., Bosisio, A., Guastamacchia, S., & Riccardi, M. (2021) 'Money laundering and the detection of bad entities: a machine learning approach for the risk assessment of anomalous ownership structures', *2020 Empirical AML Research Conference proceedings – The Central Bank of the Bahamas* [Preprint].
- Knobel, A. (2021) 'Complex Ownership Structures. Addressing the Risks for Beneficial Ownership Transparency', *Tax Justice Newrok Working paper* [Preprint].
- OECD (2019) Global Forum on Transparency and Exchange of Information for Tax Purposes: The Netherlands 2019 (Second Round).
  Available at: https://www.oecd.org/tax/transparency/global-forum-on-transparency-and-exchange-of-information-for-tax-purposes-the-netherlands-2019-second-round-fdce8e7f-en.htm
- Savona, E.U. & Riccardi, M. (eds) (2017) Assessing the risk of money laundering in Europe – Final report of project IARM. Milano: Transcrime – Università Cattolica Sacro Cuore.
  Available at: www.transcrime.it/iarm
- Savona, E.U. & Riccardi, M. (2018) Mapping the risk of organised crime infiltration in European businesses – Final report of project MORE. Università Cattolica del Sacro Cuore. Milano.
- Tavares, R. (2013) *Relationship between Money Laundering, Tax Evasion and Tax Havens*. Thematic Paper on Money Laundering. Bruxelles: European Parliament – Special Committee on Organised Crime, Corruption and Money Laundering.
  Available at: http://www.europarl.europa.eu/meetdocs/2009_2014/documents/crim/dv/tavares_ml_/tavares_ml_en.pdf (Accessed: 16 March 2016)
- Tax Justice Network (2015) Financial Secrecy Index 2015 – Final results. Tax Justice Network.
  Available at: http://www.financialsecrecyindex.com/PDF/FSI-Rankings-2015.pdf
- UNODC (2020) Covid-19 Vaccines & Corruption Risks: Preventing Corruption In The Manufacture, Allocation And Distribution Of Vaccines. COVID-19 Policy Paper.
- Van der Does de Willebois, E. *et al.* (2011) The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It. The World Bank.
  Available at: http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8894-5 (Accessed: 29 October 2014).