International Journal of Industrial Organization 104 (2026) 103228

Contents lists available at ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

# Public procurement cartels: A large-sample testing of screens using machine learning

Mihály Fazekas [a,f,*], Bence Tóth [b,f], Johannes Wachs [c,d,e], Aly Abdou [f]

[a] Central European University, Vienna, Austria
[b] University College London, London, United Kingdom
[c] Corvinus University of Budapest, Hungary
[d] ELTE Centre for Economic and Regional Studies, Budapest, Hungary
[e] Complexity Science Hub, Vienna, Austria
[f] Government Transparency Institute, Budapest, Hungary

ARTICLE INFO

Dataset link: https://zenodo.org/records/17595875

*JEL codes:*
C21
C45
C52
D22
D40
K42
L41

*Keywords:*
Cartel screening
Bid-rigging
Public procurement
Europe
Machine learning

ABSTRACT

Due to the high budgetary costs of public procurement cartels, it is crucial to measure and understand them. The literature developed screens that work well for selected cartel types and with high quality data, but it didn't produce generalisable knowledge supporting policy and law enforcement on typically available datasets. We simultaneously measure multiple cartel behaviours on publicly available data of 73 cartels from 7 European countries covering 2004–2021. We apply machine learning methods, using diverse cartel screens characterising pricing and bidding behaviours in a predictive model. Combining many indicators in a random forest algorithm achieves 70–84 % prediction accuracy, distinguishing behavioural traces of confirmed cartels from non-cartels across different cartel types and countries (accuracy is 97 % when trained and tested on a single cartel case, typical of the literature). Most screens contribute to prediction in line with theory. These results could improve cartel detection and investigations and support pro-competition policies.

## 1. Introduction

Public procurement, that is governments buying goods and services, accounts for about 12 percent of global GDP or 11 trillion USD per year (Bosio et al., 2022). Collusion among bidding firms in these markets represent a major problem, as it is more likely to arise and operate for longer periods in public procurement than in traditional markets (World Bank, 2009, 2011). Given the large volumes of spending, even a small percentage increase in prices translates into substantial budgetary costs and welfare losses. Massive amounts of data are generated daily by e-procurement systems which present great opportunities for improving both investigations and

* Corresponding author at: Central European University, Vienna, Austria & Government Transparency Institute, Budapest, Hungary: Quellenstrasse 51, Vienna 1100, Austria.

*E-mail address:* FazekasM@ceu.edu (M. Fazekas).

https://doi.org/10.1016/j.ijindorg.2025.103228
Received 17 May 2024; Received in revised form 17 November 2025; Accepted 21 November 2025
Available online 25 November 2025
0167-7187/

prevention. Recognising these opportunities, an academic literature emerged proposing risk indicators which signal potential cartel behaviour, i.e. cartel screens; and several data-driven initiatives by competition authorities have sprung up.

In spite of considerable efforts, current research has not produced comprehensive, policy applicable cartel screen tools. First, most datasets used by the academic literature include variables not widely available in e-procurement systems and tend to be high-quality compared to standard large-scale public procurement datasets. For example, academic studies include independently sourced cost estimates to quantify markups (e.g. Abrantes-Metz et al., 2006) or have low missing rates on key variables such as contract values. Second, most academic research aims at identifying one specific cartel behaviour isolated from others. For example, some studies look at cartels using bid rotation to rig a series of tenders without being noticed (Kawai et al., 2022). This focus produces screens with strong internal validity, but external validity remains problematic in the presence of unknown and diverse types of cartel behaviours as different cartel strategies give opposing signals. Third, studies consider few cartel cases to serve as a basis for generalizing to many cartel types, markets and countries. To the best of our knowledge, the most comprehensive study to date considers one cartel strategy, one sector, five countries and seven proven cases (Rodriguez et al., 2022) .[1] Fourth, a number of studies apply collusion risk indicators (or screens) knowing the exact products of the collusive market with precise market boundaries (e.g. Ishii, 2009). However, for large-scale cartel risk estimation, precisely estimating market boundaries necessary for applying a host of indicators such as market concentration is either imprecise or impractical due to resource needs.

Recognizing the major advances in the literature, while also considering the above gaps, we develop a general model for cartel detection and test its validity using a wide range of cartel risk indicators. Our core insight is that cartels leave detectable traces of collusive behaviours, and that even if they are careful, the combination of multiple weak signals of collusion can reveal them. We hope to advance both the academic debate and support policy applications by addressing each of the 4 limitations listed above. We use readily available large-scale public procurement datasets accessible in public repositories and government websites without laborious manual data collection and correction. The source public procurement datasets we use can be accessed at our regularly updated repository: https://opentender.eu/all/download (for a methodological discussion see Fazekas et al., 2024). We combine a wide variety of cartel risk indicators used in the literature which indicate many different cartel behaviours. We train machine learning algorithms on many proven cartel cases from multiple markets in multiple countries, getting us closer to a general model with high external validity. Finally, we only use screens which are defined on the contract or company levels, avoiding the pitfalls of having to define market boundaries a priori. The final matched dataset used in this analysis, the replication R scripts, and replication guidance are published on Zenodo and can be accessed at: https://zenodo.org/records/17595875.

We evaluate the accuracy of many cartel screens to identify potential cartels in large-scale public procurement data from Bulgaria, France, Hungary, Latvia, Portugal, Spain and Sweden from the 2004–2021 period. We combine this administrative data with judicial records on proven cartels from official government sources. We test cartel screens by comparing proven cartel contracts versus likely non-cartel contracts (same market after the cartel has been busted). Moreover, we compare regression, random forest, boosting, and ensemble algorithms to select the model with the highest prediction accuracy. The best model is selected based on prediction accuracy on the test set (i.e. random sample of proven cartel and non-cartel cases not used for estimating the model).

We reiterate that no single indicator or group of indicators performs well when tested against a variety of proven cases as different screens are better suited to detect some types of cartels, while missing others. Nevertheless, as each indicator is valuable to detect at least some cartels they can be combined as complements to improve prediction accuracy. Crucially for our claim about a general cartel prediction model, when a group of screens are considered together, their combined accuracy is high across all proven cases. Our best random forest model achieves 70–84 % prediction accuracy across all countries, markets, and proven cases. For the most impactful individual indicators, we identify an impact function consistent with theory despite high degrees of flexibility, non-linearity and interactions in random forest models.

We contribute to the academic literature, first, by confirming the validity of many cartel screens from prior research, especially bidding patterns-related indicators such as the number of bidders or subcontracting which are less often tested. We also show that price-based indicators, receiving most academic attention, are impactful yet not indispensable ingredients for a high accuracy cartel detection model. Second, we also establish the limits of single-indicator, single-cartel type approaches. When cartels rig tenders in diverse ways - sometimes increasing the number of bidders, sometimes decreasing it -, combining different indicators in a flexible model leads to the most accurate detection of collusion. Third, we explore the virtues and pitfalls of adding new countries and cartels to the learning dataset. We demonstrate that adding cartel cases to the model from new countries typically improves prediction accuracy for a given country, even though the improvement is uneven reflecting dissimilarities among countries. Fourth, our models achieve considerably higher prediction accuracy on a single case than on many cartels from many countries: 94 % versus 84 %. This drop in accuracy highlights the challenges of detecting multiple cartel behaviours compared to a single, well-defined case; and confirms the findings from the widest study to date, Rodriguez et al. (2022) which also finds 82–86 % accuracy rates when tested on a multi-country, multi-cartel set (versus single case accuracies of 81–95 %). Fifth, we unpack our black-box machine learning models and explore the relationships identified. This reveals that machine learning can learn theoretically sound and broadly interpretable relationships between various cartel screens and predicted cartel probability.

We contribute to policy debates, first, by pointing out data quality issues limiting prediction accuracy and theory testing, even in high-income European countries with a strong track record of e-government. Several key variables for cartel screening are not collected systematically across a wide set of countries (e.g. information of bid prices of losing bidders). Second, we confirm that

[1] Similarly comprehensive are studies of Huber et al (2022) and Huber & Imhof (2019) which look at one cartel strategy, one sector, two countries and five proven cases. These represent an approach closest to ours.

reasonably precise large-scale cartel risk estimation is possible even using such imperfect, yet readily available datasets. We demonstrate this point by predicting the risk of collusion to about 3.3 million public procurement contracts in the 7 countries in question. Our predicted scores reveal market segments with high collusion risks which could serve as a starting point for competition authorities to investigate. This is especially important given the current reliance on whistleblowers in this area.

The rest of the article is organized as follows: first, we set out the conceptual framework discussing multiple cartel types. Second, we spell out our empirical strategy, including the data used, the cartel risk indicators tested, and the methodology applied for indicator testing and prediction. Third, we compare different approaches and unpack the most accurate model. Fourth, we show the prediction of cartel risk scores for complete procurement markets and discuss policy implications. Finally, we discuss our findings and suggest future improvements.

## 2. Conceptual framework

Economists have been interested in detecting signals of collusion in markets using data on firms, bids, and prices since the 1950s with a considerable uptick in research since the 1990s (Funderburk 1974; Hewitt, McClave and Sibley 1996; Lanzillotti 1996; Porter and Zona 1999; Pesendorfer 2000; Scott 2000). These covered various markets ranging from road construction to school milk procurement (e.g. Porter and Zona 1999, Bajari and Ye 2003, Ishii 2008). Many of these works can present convincing evidence of anti-competitive behaviour because prices and costs in these markets are highly comparable, and evidence of collusion exists thanks to whistleblowers and investigations. The scope and scale of markets studied in papers on collusion has grown in recent years (Kawai and Nakabayashi 2022; Chassang and Ortner 2019; Conley and Decarolis 2016). Reliable price and bid data also enable the application of novel methods from network science (Morselli and Ouellet 2018, Wachs and Kertesz 2019, Lyra et al., 2022) and machine learning (Vadász et al. 2016; Schwalbe 2018; Huber and Imhof 2019). Given our goal of developing a general detection model of collusion, this section catalogues the diverse collusion types which can be used to guide measurement.

### 2.1. *Collusive strategies*

Collusion in public procurement aims to coordinate companies' decisions regarding price, quantity, quality or geographical presence to eliminate competition. This strategy can only be sustained if a) companies can coordinate; it is b) internally sustainable (credible punishment system, effective detection of cheating), c) externally sustainable (ability to exclude new market entrants); and d) the scheme can go undetected (i.e. no fines).

We follow the categorization of procurement collusion schemes introduced in Tóth et al. (2014) and Fazekas and Tóth (2016) which map the most important choices faced by colluding firms. We identify three dimensions: a) elementary collusion techniques, b) forms of rent-sharing, and c) resulting market structure. Elementary collusion techniques describe companies' bidding behaviour that ensures that contracts are won by the agreed supplier. These are a) withheld bids, b) non-competitive bidding, and c) joint bidding. First, companies can withhold their bids to reduce competitive pressure on the other companies and eventually raise contract prices. Second, companies can mimic competition by either submitting deliberately losing bids at inflated bid prices or erroneous bids.[2] This is the most common form of public procurement collusion, according to expert practitioners (OECD, 2014). Third, companies can submit a joint bid, which can be a sign of a special collusion scheme that also establishes the method of rent allocation.

The second dimension of collusion schemes is their rent allocation mechanism. This choice is strongly influenced by whether companies are active or passive participants in public tenders (Pesendorfer, 2000). Rent allocation is straightforward for active members of a scheme. For example, a consortium can easily formalize rent allocation through their contracts. Companies can also agree to allocate geographical markets or to win contracts cyclically, which makes rent allocation straightforward. However, if companies are not participating in tenders directly, rent reallocation must happen in alternative ways. For example, they might subcontract each other or give each other informal side-payments.

The third dimension is the market structure that follows from the various collusive strategies. First, coordination can lead to highly concentrated market structures. For example, if collusion involves splitting markets geographically or by product markets, then companies will end up with a very high share of contracts at a regional or sub-sectoral level.[3] Coordination can also lead to high market shares when passive participants get paid in alternative ways (Levenstein and Suslow, 2006; Pesendorfer, 2000) – few companies winning all contracts whereas smaller ones 'get paid' through sub-contracts or side-payments. Second, prior research also suggests that colluding suppliers can effectively imitate competitive market structure (Athey, Bagwell, and Sanchirico, 2004; Mena-Labarthe, 2012; Pesendorfer, 2000; World Bank, 2011). The cartel uses time to evade competition, with individual companies deferring profits and waiting their turn. Companies winning cyclically will not face competitive pressure and their market share will not show any timely changes.

[2] Public tenders very often award companies based on a combination of price and quality. Therefore, losing bids might just offer significantly lower quality at the same price as the winning supplier agreed to by the collusive scheme in practice.

[3] Note, that splitting geographical- or product- submarkets will not have detectable signs in the 'higher-level' market shares, that would be relevant in a competitive set-up. For example, a road construction market of a country with two big regions and several companies will look competitive if we look at market shares at the country level. If they start to collude and split the contracts so that half of the companies win all contracts from one region and the other half of the companies from the other, then we would observe an increase in market concentration in the regional sub-markets. However, when looking at the country-level picture, the market shares would be unchanged.

Each combination of a) elementary collusion techniques, b) rent allocation mechanisms, and the c) resulting market structures forms a distinct collusion strategy (Table 1).[4] As strategies vary by these measurable dimensions, we can combine (group) indicators by these theoretical scenarios. Note that whereas strategies even within the same cartel can change, many contract level dimensions are exclusionary. For example, we cannot observe single-bidding and extreme bid price ranges at the same time.

As an example, we discuss collusion type B. The main feature of this strategy is that companies submit losing bids (or they might withdraw them or submit false bids), while they share rents through subcontracts, which leads to a concentrated market structure.[5] First, there is no clear theoretical expectation on the number of submitted bids or probability of single bidding. Second, as many of the bids have to be losing bids, we expect either a) the number of withdrawn bids or faulty bids to increase, or b) bid prices to show an extreme distribution. Bid prices might be both very closely aligned together or dispersed. Third, other traces of coordinated bidding that are harder to be found in an automated way - such as identical mistakes or having the same author of the bidding documents - are also expected to occur. Fourth, if subcontracting is indeed the dominant rent-reallocation mechanism, then public procurement data might have traces of it in terms of increased probability of subcontracted contracts.[6] Alternatively, these contractual or informal relationships are outside the procurement domain and so traces of exchanges should come from alternative sources. Fifth, procurement spending should become concentrated with a few companies having high market shares.

## 3. Empirical strategy

We argue that most approaches proposed by the literature, though valuable, may overfit to specific cases. A cornerstone of modern machine learning practice is the evaluation of predictive algorithms on unseen data. Perhaps owing to the rarity of clean data on proven cartel cases, few research papers apply cartel screening methods to multiple examples (for a recent exception see Huber et al., 2022). We see an opportunity to widen the scope of cartel screening to large, heterogeneous markets with varying data quality by ensembling, or combining, multiple cartel screens (extending earlier attempts by Tóth et al., 2014). We borrow the term ensembling from the machine learning literature, which has long recognized that combining weaker signals can produce a much stronger predictive model (Breiman 2001). Below, we examine whether such an approach can overcome the challenges of noisy data and heterogeneous markets and cartel behaviours to produce accurate methods of detecting cartels and anti-competitive behaviours.

### *3.1. Data*

This section presents the data used in the analysis. We give a high-level overview of the procurement data, outline our search for data on cartel cases, and explain how we linked proven cartel cases with public contracting datasets. We present summary statistics by country that show the number of matching contracts used for model estimation. More details on data and data processing are in Appendix A.

### *3.2. Selection based on procurement data*

As a first step, we screened several countries with sufficient procurement data quality for the analysis. We used two main criteria for shortlisting: availability and quality of public procurement data (especially the availability of bidder information) and the number of cartel cases that overlap with procurement data. Based on this screening process and additional data quality checks - e.g. availability of long enough historical data and availability of key variables,[7] we selected 7 countries: Bulgaria, France, Hungary, Latvia, Portugal, Spain, and Sweden.

The contract-level public procurement data used is collected from official government publication websites and open data repositories. This data is therefore continuously available, and our models can be readily applied to continuous monitoring of procurement markets. For all analysed countries - except Sweden - we use data collected and published by opentender.eu.[8] It compiles contract-level data on European public procurement, covering both above- and below-EU-threshold contracts. A broader dataset, following the same structure and standardization steps, includes contracts data for 42 countries from around the world (Fazekas et al., 2024). It encompasses over 72 million contracts from 2006 to 2021.

Given the inconsistencies in data publication formats across countries, we standardized the collected information to align with a

[4] Note, that not every combination is conceptually meaningful, while some dimensions are not possible to measure with indicators based on public procurement or company data (e.g. informal side-payments are hard to observe).

[5] Note that we would see concentrated market structure based on the share of public contracts won and not necessarily based on the turnover of the participating companies due to subcontracting.

[6] Most procurement systems collect information on whether a particular supplier won a contract with explicit mention of subcontracting parts of it.

[7] For example, we analysed the availability of company names, bidder numbers, dates, that are all key for the analysis.

[8] The full public procurement data can be downloaded on http://opentender.eu/download. For a technical explanation of the database building process, see: https://github.com/digiwhist/wp2_documents/blob/master/d2_8.pdf. Note, we use data provided by Visma Opic for analyzing the Swedish proven cartel cases, that was used by the authors in Fazekas and Tóth (2016).

**Table 1**
Main characteristics of collusion types and the availability of indicators.

| Resulting market structure | Elementary collusion technique | Form of rent sharing | | | |
|---|---|---|---|---|---|
| | | Sub-contractor | Consortia/ joint ownership | Coordinated bidding | Informal side-payments |
| Concentrated market structure | Withheld bids | A | | | |
| | Losing bids | B | | | |
| | Joint bids | | C | | |
| Stable market structure | Withheld bids | D | | F | |
| | Losing bids | E | | G | |
| | Joint bids | | | | |

Notes: This table outlines the main characteristics of different types of collusion, categorized by the resulting market structure, elementary collusion techniques, and forms of rent sharing. Each cell (labeled A–G) represents a specific collusive strategy. The columns indicate the form of rent-sharing, such as subcontracting, joint ownership or consortia, coordinated bidding, and informal side-payments.
The table distinguishes between theoretically feasible and infeasible cartel strategies: Green cells: All three dimensions can be captured with available data. Orange cell: the form of rent sharing is not captured by administrative data. Grey cells: The combination does not occur in theory.
Source: Fazekas and Toth (2016).

common data structure.[9] For each country, we record and standardize key procurement details, including buyer and supplier information (names and addresses), organization location, product codes, final and estimated contract values, bid-level data (when available), and details of the contracting process such as contract award dates and procedure types. Additionally, we have implemented several data quality enhancements to improve variables critical to the analysis. These include generating unique organization identifiers to reconcile different representations of the same entity, correcting missing company identifiers and contract values, and filtering out irrelevant data points, such as direct contracts where coordination is unlikely or infeasible.

### *3.3. Cartel case collection*

We collected information on proven cartel cases manually from country specific sources of court rulings by following three search strategies: a) searching competition authorities' online repositories of proven cases, b) screening competition authorities and courts annual reports, c) contacting competition authorities of selected counties and requesting an overview of proven cases (see Appendix A for more details).

We collected cartel-level information manually. We then stored the relevant details in a standardized data template for each case.[10] The key information extracted in this process included: the names of the companies involved, the public authority that conducted the public procurement process(es) in question, the time period in which the cartel operated,[11] information related to the relevant public tender(s) (e.g. tender IDs, product types), and the location.

### *3.4. Data linking*

As explained above, we set out to collect the most important dimensions of each proven case into a structured dataset and match on as many dimensions as possible. However, matching cases based on all available dimensions proved hard and impractical. Most often the number of contracts that were awarded to one of the cartel companies, were awarded or advertised during the cartel period and managed by a public buyer that is explicitly mentioned in the court rulings were very small and often zero. Therefore, we had to apply a more lenient approach and only match by company names and the proven cartel time period that are explicitly mentioned in the cartel documents.[12] While identifying all rigged contracts unambiguously would be clearly important to find statistically meaningful patterns in the indicators, we also accept that the number of truly rigged contracts vs. the ones that could be proven at the court can (sometimes considerably) differ. We only consider the period during which the cartel was functioning (based on the court records) as collusive in our main models. We exclude contracts from before the official start date of the cartel based on the court records in the main analysis as start dates might be less reliable (and sometimes the information is missing from the court records). In Appendix C we show that our results are largely the same on samples which also consider the period before the official start dates of the cartels as non-cartel contracts, whenever given in the court records.

[9] An overview of the data structure and standardization process is available at: https://www.govtransparency.eu/government-transparency-institute-2025-public-procurement-data-processing-version-1-0/

[10] We used the Google Document translator to understand the case documents in countries where the source files were only available in the national language (e.g. Swedish, Latvian).

[11] The case documents contained varying detail on the start and end date of the cartels. In some cases, precise dates or the months of the start and end of cartel activity were defined (e.g. in most of the Swedish, and Latvian cases, and some of the French cases). In other cases, only the years were given (e.g. in most of the Spanish and Portuguese cases, and some of the French, and Latvian cases), hence we simply marked the full year, i.e. 12 months, as an estimation of the cartels' length, which might overestimate the length of the cartel activity.

[12] Note, that in some cases (for example, for all Hungarian cases), the case data was often not clear enough for assigning monthly values, hence we marked whole years that were mentioned in the court documents.

### 3.5. Final dataset

As Table 2 shows, we have 73 cartel cases in total from the seven analysed countries. We have 9068 contracts won by cartel members after the collusive period and 6548 contracts during the cartel period. Note that the number of contracts used for testing varies by cartel. For extrapolation, we use all available contracts – around 3.3 million across the seven analysed countries.

### 3.6. Indicators

Table 3 lists all cartel screens implemented in our models, along with concise definitions.[13] The indicator groups capture different outcomes of anti-competitive behaviours: i) measuring how prices are consistently at odds with competitive pricing; or ii) capturing bidding patterns that show how companies strategically lose (or unrealistically win) contracts in certain markets. In the final analysis, we could only include those variables that are consistently available across countries and allow for cartel risk predictions without market definitions.[14]

Some indicators are defined at the contract level, while others are at the company-year level. We use aggregations of contract-level indicators to the company year level to provide an additional lens on collusion risk. This approach captures persistent behavioural patterns that may not be evident in single transactions, allowing for a more comprehensive assessment of long-term, stable collusion. Table 4 presents descriptive statistics of the indicators used in the labelled data, with further details available in Appendix B. Most indicators have relatively low missing rates. The higher missing rate for the Relative Price indicator is primarily due to the non-publication of reference prices in several countries - for example, this indicator is entirely missing from the Swedish data, and has high missing rates in France and Spain. Similarly, the Benford's Law (by buyer-year) indicator has a higher share of missing values because it requires at least 100 contracts per buyer-year to allow for a meaningful distributional comparison. For buyers with fewer contracts, the indicator cannot be reliably computed.

Collusion tends to persist over time, and the creation as well as the demise of cartels is likely to reveal notable changes in competitive behaviour (Toth et al., 2014). Therefore, we include the one-year lagged versions of all company-year variables in the models. We replaced missing values in continuous variables with the average value of the variable in the labelled data and with an additional missing value for categorical variables.

### 3.7. Relative price

Relative price is defined as the final contract price divided by its initial estimate.[15] Healthy competition ought to lead to lower prices (i.e. bigger discounts) compared to the initial estimate - hence relative price can proxy competition. As collusion is about generating rents - either through higher prices or lower quality - an increase in relative contract values can be a by-product of bid-rigging schemes.

However, relative price can be affected by a range of factors. Public buyers might lack the capacity to assess market prices accurately (i.e. there is noise in the estimated price), they might be incentivised to underestimate their tender prices for administrative reasons (e.g. bad incentives for budgetary planning). Furthermore, other anti-competitive practices, such as corruption, also can bias the estimated prices: buyers might deliberately overestimate prices to avoid clear signs of overpricing. All of these factors affect both what we find in the collusive tenders, and of course in the control tenders. For example, if half of the market is captured by a cartel, but the comparison group is corrupt, then we do not expect to find significant relative price differences between these contract groups.

With the limitations kept in mind, the literature on collusion also uses prices, and relative prices in particular, to analyse bid rigging schemes. Odd price increases that cannot be explained by costs as well as long term price stability at unusually high levels indicate market performance problems OECD (2014) and Oxera (2013). Prior research has found that tenders with large discounts (relative price below 90 %) have a significant relationship with the number, capacity and experience of bidding suppliers, whereas these dimensions are unrelated to prices if discounts are small (relative price is above 90 %) (Morozov and Podkolzina, 2013). Others have used relative winning price (in combination with low bid price variance) to distinguish between collusive vs. competitive tenders for modelling favour exchanges among bid-rigging suppliers (Ishii, 2009).

### 3.8. Benford's law

Benford's Law, or the First-Digit Law, states that in many naturally occurring datasets, the first digits of numbers are more likely to

[13] See Table A.2 in Appendix B for indicator availability across countries in the labelled data. Note that for the multivariate models we also use company size, that is the number of contracts won by company in a year, and sector, that is the 2-digit CPV code, as control variables.

[14] Indicators defined on the market level require a precise definition of markets matching to the boundaries set by the cartel itself which is difficult to reliably calculate across the whole procurement sector. There are simplistic, hence likely biased approaches to define markets. For example, one is based on procurement classifications such as CPV (Common Procurement Vocabulary) and NUTS (Nomenclature of Territorial Units for Statistics) (Fazekas and Tóth, 2016).

[15] Note that the rules differ between countries on initial estimate calculation. For example, it can be an average estimation of market prices, but also an upper-bound estimation so that public buyers choose a more competitive procedure type. The higher the estimated tender value is, the more competitive (or at least more regulated) the tendering procedure is in most regulatory regimes.

**Table 2**
Analysis dataset scope by country.

| COUNTRY | NUMBER OF CARTELS | LABELLED DATA (NUMBER OF CONTRACTS) | | EXTRAPOLATION DATA | |
|---|---|---|---|---|---|
| | | DURING CARTEL PERIOD | AFTER CARTEL PERIOD | TOTAL NUMBER OF CONTRACTS | YEARS |
| BG | 2 | 187 | 141 | 219,454 | 2010–2021 |
| ES | 15 | 4681 | 3179 | 352,582 | 2005–2019 |
| FR | 10 | 178 | 1619 | 2195,479 | 2004–2018 |
| HU | 18 | 664 | 742 | 79,550 | 2005–2012 |
| LV | 20 | 524 | 2784 | 144,835 | 2006–2020 |
| PT | 2 | 56 | 106 | 62,445 | 2009–2018 |
| SE | 6 | 258 | 497 | 244,839 | 2009–2018 |
| **Total** | **73** | **6548** | **9068** | **3299,184** | |

Notes: This table summarizes the analysis dataset by country. For each country, we report the number of cartels identified from court documents, the number of labelled contracts during and after the cartel period. The final columns indicate the number of contracts included in the extrapolation dataset and the time period covered.

be small, with the digit '1′ appearing most frequently. In the context of procurement, significant deviations from this expected distribution can indicate potential bid-rigging or other fraudulent activities, as markets or buyers engaged in collusive behaviour may exhibit patterns that diverge from the typical distribution. By applying Benford's Law to bid prices across contracts, we can identify abnormal patterns. Similar cartel detection screens have been used by Samà (2014) to detect collusion in the London Metal Exchange. Specifically, we compute the mean absolute deviation of the yearly distribution of leading digits in bid prices within each buyer's and market's contracts and compare them to the expected frequency distribution as predicted by Benford's Law.[16]

### 3.9. Single bidding and number of bidders

Withholding bids is one of the most straightforward ways to rig a tender, which results in a low number of bids and higher probability of single-bidder contracts by definition. While empirical research focusing on single-bidding as a collusion indicator is slim (Barrus 2011; Tóth et al., 2014), competition policy guidelines cite it as one possible elementary technique (OECD 2014; SCA 2015). Submitting fake bids is time-consuming, costly, and poses its own risks for the cartel members (for instance if the same language is used in multiple bids or if such an effort requires additional communication and coordination).

The reliability of bidder number based collusion indicators – especially single bidding - are affected by three possible confounding factors. First, it is a bluntly obvious signal of anti-competitive risks, that collusive companies might want to hide - especially if they are participating in markets with historically many bidders. Second, single-bidding is also a potential side-effect of corruption in public contracting, as favouring well-connected suppliers can exclude outsider companies entirely from the bidding process (Fazekas et al., 2018). While the first issue suggests that many cartels could operate on a basis that cannot be captured by this indicator, the second warns us about the limitations of this indicator used for indicating collusion specifically - as it might capture other anti-competitive behaviours instead. Third, a market with many single-bidder contracts attracts not only the attention of the competition and anti-corruption authorities, but also of potential competitors who would naturally see such a market and its high markups as a target for expansion, making the sustainability of the collusive agreement less viable.

Nevertheless, some of the bid-rigging schemes can be picked up even by this simple indicator. One relative strength of this approach to collusion is that it is easy to organize. Indeed, Barrus' (2011) study of the Kentucky highway construction market links single-bid contracts to tacit collusion. Such behaviour may be highly visible, but it is difficult to legally prove that it is the result of illegal coordination.

### 3.10. Number of buyers and markets – withheld bids

We use the number of unique buyers and markets a company engages with each year as indicators of potential withheld bids. Similar to the previously discussed indicators based on the number of submitted bids, these indicators reflect how colluding companies may strategically withhold their bids from specific tenders or markets. Companies can withhold their bids by different market dimensions that can make coordination easier, hence lower the costs of cartel maintenance: from specific sub-markets (e.g. based on CPV codes), from specific (group of) buyers, or geographical location. Such techniques are also commonly quoted as a possible cartel strategy (OECD 2014; SCA 2015).

In our analysis we calculate two versions of the missing bidder indicators: a) number of unique buyers a supplier wins a contract from, b) number of unique markets a supplier wins a contract on. To ensure comparability across suppliers of different sizes, these values are normalized by the total number of contracts a supplier wins per year. For example, in the bivariate tests, we expect that companies win contracts from more buyers and more markets (i.e. measured as the number of unique product codes assigned to their contracts) after the cartel period has ended.

[16] Based on theoretical considerations, we calculate the indicator only for buyers with >100 contracts in the complete procurement data (both labelled and unlabelled contracts). This ensures sufficient variation in contract values to meaningfully observe Benford's Law.

**Table 3**
Cartel risk indicators used in the analysis.

| Category | Nr | Indicator | Lowest level of observation | Is aggregated | Description |
|---|---|---|---|---|---|
| **Prices** | 1 | Relative price | Contract | Yes | Ratio of the final price and the estimated price |
| | 2.1 | Benford's law | Market-period | No | The mean absolute deviation of contract prices within a CPV-3 digit market from the expected distribution of leading digits as described by Benford's Law. |
| | 2.2 | Benford's law | Buyer-period | No | The mean absolute deviation of a buyer's contract prices from the expected distribution of leading digits as described by Benford's Law |
| **Bidding patterns** | 3.1 | Single bidding | Company-year | Yes | Contract receiving a single bid during the tendering process. |
| | 3.2 | Number of bids | Contract | Yes | Number of bids received per contract |
| | 4.1 | Number of buyers | Company-year | No | The number of unique buyer companies to which a company submits bids for contracts |
| | 4.2 | Number of markets | Company-year | No | The number of unique markets to which a company submits bids for contracts* |
| | 5 | Subcontracting | Contract | Yes | Whether a contract has a subcontractor. |
| | 6 | Consortia | Contract | No** | Whether the winning bid was a consortium. |

Notes: Cartel risk indicators by category, including their level of observation, aggregation status, and short descriptions. Product markets are defined using 3-digit CPV codes. Consortia cannot be aggregated to the company level as only the consortium name is available.

**Table 4**
Descriptive Statistics of Key Cartel Risk Indicators in the Labelled Dataset (All Seven Countries).

| Variable | Missing rate/ contracts | Average | Standard Deviation | Min. | Median | Max. | Histogram |
|---|---|---|---|---|---|---|---|
| Number of bids | 0.20 | 5.56 | 7.05 | 0.00 | 3.00 | 50 | |
| Number of bids (aggregated) | 0.16 | 5.71 | 4.65 | 1.00 | 4.03 | 29.5 | |
| Number of bids (aggregated lagged) | 0.19 | 6.12 | 5.05 | 1.00 | 4.09 | 35.3 | |
| Single bidding (aggregated) | 0.16 | 0.25 | 0.25 | 0.00 | 0.17 | 1.00 | |
| Single bidding (aggregated lagged) | 0.19 | 0.25 | 0.26 | 0.00 | 0.16 | 1.00 | |
| Subcontracting (aggregated) | 0.30 | 0.09 | 0.13 | 0.00 | 0.05 | 1.00 | |
| Subcontracting (aggregated lagged) | 0.32 | 0.10 | 0.14 | 0.00 | 0.05 | 1.00 | |
| Relative price | 0.71 | 0.92 | 0.15 | 0.50 | 1.00 | 1.29 | |
| Relative price (aggregated) | 0.27 | 0.91 | 0.12 | 0.50 | 0.93 | 1.26 | |
| Relative price (aggregated lagged) | 0.33 | 0.91 | 0.12 | 0.50 | 0.93 | 1.26 | |
| Number of markets (aggregated) | 0.10 | 0.27 | 0.23 | 0.02 | 0.23 | 1.00 | |
| Number of markets (aggregated lagged) | 0.17 | 0.28 | 0.23 | 0.02 | 0.24 | 1.00 | |
| Number of buyers (aggregated) | 0.09 | 0.58 | 0.27 | 0.04 | 0.62 | 1.00 | |
| Number of buyers (aggregated lagged) | 0.17 | 0.61 | 0.28 | 0.04 | 0.67 | 1.00 | |
| Company size (number of contracts) | 0.00 | 81.3 | 126.63 | 0.00 | 29.00 | 538 | |
| Benford's Law by buyer/year | 0.78 | 0.03 | 0.02 | 0.01 | 0.03 | 0.19 | |
| Benford's Law by buyer/year (aggregated) | 0.36 | 0.03 | 0.01 | 0.01 | 0.03 | 0.14 | |
| Benford's Law by buyer/year (aggregated lagged) | 0.40 | 0.03 | 0.01 | 0.01 | 0.03 | 0.14 | |
| Benford's Law by market/year | 0.35 | 0.02 | 0.01 | 0.00 | 0.02 | 0.17 | |
| Benford's Law by market/year (aggregated) | 0.20 | 0.02 | 0.01 | 0.00 | 0.02 | 0.10 | |
| Benford's Law by market/year (aggregated lagged) | 0.25 | 0.02 | 0.01 | 0.00 | 0.02 | 0.11 | |
| Subcontracting | 0.31 | 0.09 | 0.29 | 0.00 | 0.00 | 1 | |
| Consortia | 0.00 | 0.09 | 0.29 | 0.00 | 0.00 | 1 | |

Notes: Descriptive statistics for key cartel risk indicators for the labelled data across all seven countries. The table reports missing rates, averages, standard deviations, and distribution characteristics for both raw and aggregated indicators. Histograms illustrate the distribution shape. Aggregated and lagged versions are computed at the company-year level where applicable.

### *3.11. Consortia*

Instead of withholding bids – captured by the previous indicators – companies can also decide to submit joint bids, that is another elementary collusion technique. By joint bidding, companies lower competition and facilitate communication therefore it can be used as a price-fixing tool (Albano et al., 2009). Joint bidding also acts as an enforcement mechanism, as rent sharing is agreed in a formal contract. Due to the nature of procurement data, calculating the consortium indicator requires additional data processing,[17] and once we had a good indication of consortia status, we connected the already known cartel member names based on a simple string matching to decide whether cartel members were also part of the joint bidding.

### *3.12. Subcontracting*

Rent division between cartel members is a challenge (Asker, 2010), as transferring money between cartel members is risky - receiving money from a competitor is a signal of potential collusion. Another simple way to reallocate rents is through subcontracts. The prevalence of subcontracting in public contracts is contentious. While it can increase competition and efficiency through cooperation and knowledge exchange (Albano, Spagnolo, and Zanza, 2009; Estache and Limi, 2008), it can signal a collusive arrangement and serve as a tool for rent-reallocation (Carbone et al., 2024; Fazekas and Tóth, 2016; Alexander, 1997). Therefore, while subcontracting in itself – as neither of the indicators - is not a strong enough indicator of collusion, it can indicate a form of rent sharing if other red flags are also present. We analyse subcontracting at the contract-level and calculate the share of contracts using subcontracting in all collusive vs. competitive contracts.[18]

### *3.13. Methods*

This section outlines how we test collusion risk indicators by exploiting the differences between proven cartel and likely non-cartel contracts. We test all individual indicators in competing multivariate models such as logistic regression, random forests, gradient boosting machines, and ensemble methods. We select the best model based on test-set accuracy.

Our models predict the binary cartel/non-cartel outcome variable. Contracts are flagged as collusive when they have been awarded to convicted companies during the proven cartel period, irrespective of the particular product group, that is companies are allowed to collude on a range of product markets (filled red symbols in Fig. 1). Conversely, contracts are flagged as non-collusive when they are awarded to the same companies, but after the cartel was busted (grey filled symbols). The contract grouping allows for contract-level

[17] For example, there is no clear indicator in the source data on whether a given bid is submitted by a group of companies, hence we need to find them based on an algorithm – detailed in Appendix B.
[18] Note, that the indicator could be calculated in the following countries: Bulgaria, Latvia, Spain, and France.

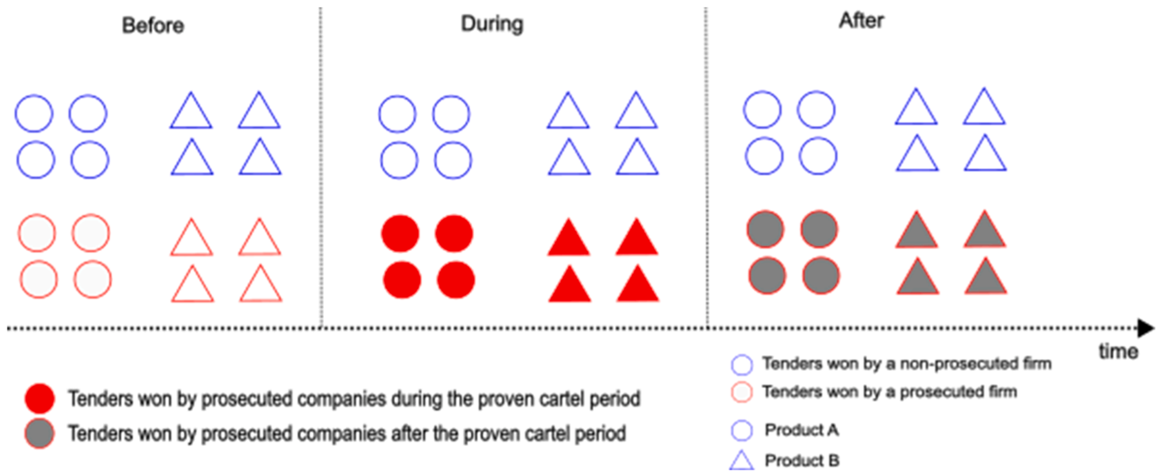


**Fig. 1.** Tender Classification Based on Cartel Involvement Over Time
Notes: This figure illustrates how tenders were labelled before, during and after the proven cartel period. Circles represent Product A and triangles represent Product B as an example, however, we are agnostic to product markets when labelling our data, we simply label all tenders won by a prosecuted firm. Red-filled shapes indicate tenders won by prosecuted firms during the cartel period and are flagged as collusive. Grey-filled shapes represent tenders awarded to the same firms after the official start of the cartel investigation, hence flagged as non-collusive (i.e. we assume cartels stop once they are investigated). Red lined, hollow shapes indicate tenders awarded before the recorded start date of the collusive behaviors. Blue outlined shapes indicate tenders awarded to non-cartel firms - i.e. these would be part of our prediction dataset. Our training data consist of red and grey tenders, and we do a robustness test using tenders awarded before the proven cartel start date by prosecuted companies as non-collusive tenders.

indicators as well as company-year level indicators based on the winning supplier's aggregate characteristics.

As cartels are diverse and public procurement datasets are generally noisy (e.g. widespread missing, erroneous data), elementary collusion risk indicators are imprecise outside of well-delineated and homogeneous collusive markets (more on this argument see Fazekas et al., 2023). The validity and reliability of a cartel risk detection framework can be increased if elementary indicators are combined into either a sequence of tests (e.g. Tóth et al., 2014) or into a composite risk score (e.g. Huber and Imhof, 2019). In the absence of sufficiently precise theory guiding the methods for combining elementary collusion risk indicators, including prescribing indicator weights, we turn to data-driven approaches. The main advantage of such approaches is that they can directly learn from the patterns in the data to identify the accurate combination of individual indicators. We compare four different methods[19] to identify the model with the highest prediction accuracy:

- Binary logistic regression;
- Random forests;
- Gradient boosting machines; and
- Ensemble methods.

Such supervised machine learning approaches make use of a wide set of labelled cartel and non-cartel cases to learn how best to predict that label in the training set, making use of a set of individual collusion risk indicators and control variables (for a similar approach see Huber et al., 2022). For model comparison, we calculate the accuracy of model prediction on unseen, so-called test dataset; that is a dataset which was not used to fit the model. Accuracy is defined as the share of correctly classified contracts over all contracts in the test set, using a method called k-fold cross validation (James et al., 2021). We use three different definitions of test datasets to reflect different types and degrees of difficulties for prediction tasks.

- Random sample of contracts (five-fold cross validation) :[20] This is considered the easiest prediction task as contracts belonging to cartel cases used to train the model are used to test prediction accuracy. It is considered a policy relevant scenario, i.e. identifying further cartel contracts for already known cartel cases, however, it is not necessarily the most typical one. This test-train split is the most comparable set-up to prior studies.

[19] We also tested neural networks with various configurations; however, they consistently underperformed compared to Random Forest, Gradient Boosting, and Ensemble models.

[20] The random sample is drawn using cluster sampling based on company-year clusters, ensuring that entire clusters are assigned to either the training or test set. This prevents data leakage, that is no contract from the same cluster appear in both sets, preventing the model from learning which specific company is colluding from aggregate company-level information.

- Random sample of cartels (five-fold cross validation): We consider this set-up as the most relevant albeit more demanding than the previous one. This is because in a typical competition policy use case, our algorithm would be used to identify completely unknown cartels based on patterns in known cases. Considering the more demanding nature of this test-train split, we increased the proportion of the training sample.
- Leave one country out (seven-fold cross validation) :[21] Given our main goal of developing a general model accurately predicting cartel behaviour, we are particularly interested in whether cartels from one country can predict cartels in another. We consider this as the most demanding test to our models, given cross-country differences in data, cartel and non-cartel behaviours.

While we compare four generic types of models, as outlined above, we tune each of these first so that the best variant of the model can be compared to the other methods.[22] Binary logistic regression models are considered as the baseline as they are the least variable model, that is they are the least sensitive to different test-train splits. However, we expect binary logit models to have the lowest accuracy too, given the more restrictive constraints the model imposes on parameters.

In the context of fraud, corruption, and collusion, various studies have used the random forest and gradient boosting machine algorithms. We also use these methods because of their ability to model a diverse array of different collusive strategies and the markers they leave. Random forests and gradient boosting machines are supervised machine learning methods which predict the output by constructing multiple decision trees with given features (Breiman, 2001). They are particularly well suited for datasets with many explanatory variables and where the same outcome may be the result of multiple different combinations of predictor values (James et al., 2021). We also use ensemble methods which combine predictions from multiple machine learning algorithms to generate a single prediction. The ensemble model combines the predictions of five individual models: Random Forest, Extreme Gradient Boosting, Support Vector Machine, Generalized Additive Model, and Conditional Inference Forest.[23] These models were selected with the aim of improving overall prediction accuracy by leveraging tree-based methods for capturing complex interactions, kernel-based approaches for handling non-linear patterns, and additive modelling for greater flexibility in feature effects. This helps us avoid the pitfalls of relying too heavily on a single algorithm and improves the robustness of our predictions. By combining multiple algorithms, we can capture a broader range of features in the data and account for different sources of variability.

Despite their flexibility and suitability for the complex prediction problem, these models lead to results which are harder to interpret. In other words, to achieve high prediction accuracy, we must sacrifice some degree of interpretability. Nevertheless, we will explore the relationships identified by the best model below to compare our theoretical expectations and best empirical results.

We run a series of robustness tests, i) on different samples (Appendix C) and ii) using different sets of predictors (Appendix D). Appendix C.1 demonstrates that including contracts of collusive firms before the official start date of the cartel as non-cartel contracts results in a largely similar model. Appendix C.2 shows results for models trained and tested on only one cartel case, which represent model variants more directly comparable to prior research. Appendix C.3 shows how adding new countries to the training set contributes to prediction accuracy for a selected country. Appendix D.1 contains models estimated using restricted sets of cartel screens excluding indicators which are determined before bidding such as product market. Appendix D.2 demonstrates that our models perform similarly well when missing values are treated differently and when continuous indicators are mean centred by country.

### *3.14. Predicting cartel risks*

Once the highest accuracy model is selected, its predictions can be extrapolated to the full set of contracts across all analysed countries, encompassing approximately 3.3 million contracts. This extrapolation is predicated on the assumption that cartel behaviours in the whole economy are comparable to the proven cases and that the underlying data points are also comparable (e.g. variable distributions, missing rates, variable availability). Given that we analyse a wide range of proven cartels and use standard public procurement datasets collected from government publication portals, we consider these preconditions met.

## 4. Results

### *4.1. Selecting the model*

This section describes four different predictive models - logistic regression, random forest, gradient boosting machines, and an ensemble model – and it compares their prediction accuracy on three different test datasets – random contracts sample, random cartel sample, and "leave one country out" sample. Before making such comparisons, we optimised each model to achieve highest prediction accuracy on the training set. The binary logit model was not tuned as it is rather inflexible. We included all individual collusion risk indicators, and the control variables described above.

[21] Please note folds equate countries in this test, implying that some test sets are larger than others.

[22] We implemented data preparations and modelling in R (version 4.1.2). For logistic models, we used stats library and glm function; for Random Forests we used the randomForest library and randomForest function; for Gradient Boosting Machines we used the gbm library and gbm function; for Ensemble methods we used the SuperLearner package and SuperLearner function.

[23] The ensemble model was implemented within the SuperLearner framework in R, utilizing models from several key packages: randomForest (SL.randomForest), kernlab (SL.ksvm for Support Vector Machines with a Radial Basis Function Kernel), mgcv (SL.gam for Generalized Additive Models), party (SL.cforest for Conditional Inference Forests), and xgboost (SL.xgboost for Extreme Gradient Boosting)

For the Random Forest model, we optimised the number of trees (optimal parameter is 1000 trees) and the number of variables used in each tree (optimal parameter is ten variables). For Gradient Boosting, the number of trees (optimal number of trees is 3000) and interaction depth (optimal depth is six), shrinkage (optimal shrinkage is 0.01), and minimum number of observations per node (optimal is ten) were optimised. For the Ensemble model, we used a combination of models, including Random Forests, Gradient Boosting, Support Vector Machines, Generalized Additive Models, and Conditional Inference Forest.

The four models perform differently on the three different test sets reflecting the different challenges of each set-up (Table 5). Considering the scenario most comparable to prior studies where we already know or suspect some cartel contracts and the corresponding colluding firms and aim to identify further rigged contracts, all four models achieve high prediction accuracy. Prediction accuracy is the highest for Random Forest (accuracy approximately 84 % and AUC around 0.90), Boosting and the Ensemble methods achieve slightly lower results (Table 5).[24]

Considering a more demanding and most policy-relevant scenario when completely unknown cartels shall be identified by the model, all four models drop in prediction accuracy, going down from about 84 % to about 72 % correctly predicted cases in both the Random Forest and the Gradient Boosting model (AUCs also drop to about 0.71 for the Gradient Boosting and to 0.66 for the Random forest). The Ensemble method exhibited marginal differences, about 71 % accuracy and AUC of 0.68. An even more demanding test to our models is when all cartels of a country are preserved for the test set, compounding both cross-cartel and cross-country variation. Model performance slightly drops further, while the best models remain 69–71 % accurate, with AUC of 0.60.

The best Random Forest model achieves 84 % accuracy, AUC of 0.90, and a False Positive Rate (FPR) of 12 %, using five-fold cross validation (prediction accuracy goes up to 93–97 % when tested on a single cartel case, typical of prior literature, see Appendix C.2). Its performance is also relatively balanced. Its precision is 80.5 % and recall is 81.1 %, that is when it predicts cartels, it is correct in 80 % of cases, while it fails to identify 18.9 % of actual cartel cases. (Table 6).

To test the robustness of our models, we evaluate the models on an alternative training dataset where contracts awarded before the proven start date of the cartels are included in the analysis as non-cartel contracts (see Appendix C.1). The Random Forest model remains largely the best model across all tests – although accuracy scores slightly drop for the sampling by contracts and cartels tests. However, including the pre cartel contracts to the training data improves the accuracy for the 'leave one country out' test, increasing the Radom Forest accuracy from 69 % to 75 %. The two best Random Forest models' predictions on the full dataset (main and robustness sample modes) are highly correlated (linear correlation coefficient=0.79) albeit the average predicted probabilities are lower for the model trained on the alternative sample, when contracts prior the proven start date are also included.

### 4.2. Unpacking the final best model

While the overall prediction performance of the model is key, from a policy perspective, it is also crucially important to open the black box of machine learning and explore the estimated response functions and their fit with theory. We consider any predictive model useful for policy uses and supporting further scientific work if it is not only accurate, but the estimated relationships between predicted cartel risks and the individual collusion risk indicators match theoretical expectations. Hence, we review the most important predictors and their impact on predicted cartel risks.

To better understand which features drive accurate detection of potential collusion, Fig. 2 presents the variable importance plot from the Random Forest model. It ranks predictors based on their contribution to classification accuracy, measured by the mean decrease in Gini impurity. Variables that lead to larger decreases in classification error are considered more important and are ranked higher in the figure.

Besides the economic sector (broadly defined product market) and company size, cartel screens, either on the contract or company-year levels, are most important for precise prediction (Fig. 2).[25] Among the individual collusion risk indicators, the most important are the number of bids (company-year aggregate in t), Benford's law compliance by market and year (company-year aggregate in t and t-1), number of buyers (company year aggregate in t), number of markets (company-year aggregate in t) (Fig. 2). Contract level predictors are moderate and less important, underlying the systematic and relational nature of collusion. Appendix D.1 showcases model performance using different sets of predictors, underlying that model without indicators prior to bidding such as product market information perform comparably to the main model (74–78 % accuracy compared to 84 % on the random sample of contracts test).

We start by analysing partial dependence plots for some of the most important collusion risk indicators to establish whether identified relationships are in line with theory (Fig. 3, panels A-D). First, for the number of bids (aggregated at the company-year level

[24] Other feature engineering approaches have been tested, such as replacing missing values with a dummy value ('−999') instead of the average across the labelled data and centering the indicators around the county average; however, the results remained unchanged. The outcomes of alternative models are presented in Table A.10 in Appendix D.2.

[25] We tested if structural elements, like product market and company size, explain a disproportionately large share of variation in the collusion status by estimating non-interacted logit models (Table A.9 in Appendix D.1). We find that the explanatory power of models based on only these two variables explain between 8-14% of the total variation, while adding cartel indicators increases the explained variance to 27%. Findings from non-interacted logistic regressions and highly interacted random forest models together highlight that predictions are most accurate when the interaction between some structural features and behavioral cartel screens are taken into account. Crucially for our empirical claims, we find that structural features on their own are not sufficient for accurate cartel prediction.

**Table 5**
Main model results.

| Model (test set) | Metrics | Logit | Random forest | Boosting | Ensemble |
|---|---|---|---|---|---|
| **Random sample of contracts (5-fold cross validation)** | Accuracy | 75.58 % | 83.96 % | 82.41 % | 81.88 % |
| | AUC | 0.73 | 0.90 | 0.90 | 0.90 |
| | FPR | 13.95 % | 11.50 % | 17.79 % | 18.27 % |
| **Random sample of cartels (5-fold cross validation)** | Accuracy | 71.01 % | 71.63 % | 72.19 % | 71.01 % |
| | AUC | 0.59 | 0.66 | 0.71 | 0.68 |
| | FPR | 37.41 % | 43.22 % | 37.78 % | 34.88 % |
| **Leave 1 country out** | Accuracy | 71.41 % | 69.18 % | 71.39 % | 70.28 % |
| | AUC | 0.54 | 0.58 | 0.60 | 0.60 |
| | FPR | 22.13 % | 19.81 % | 29.97 % | 24.12 % |

Notes: Table shows the accuracy, AUC, and false positive rates (FPR) for Logit, Random Forest, Boosting, and Ensemble models evaluated using 5-fold cross-validation on random contract samples, cartel samples, and leave-one-country-out tests.

**Table 6**
Confusion matrix, best Random Forest model, all countries Combined, Test set: 20 % of company-year pairs (~19 % of contracts).

| | | Reference | |
|---|---|---|---|
| Prediction | | No | Yes |
| | No | 1519 | 225 |
| | Yes | 234 | 964 |

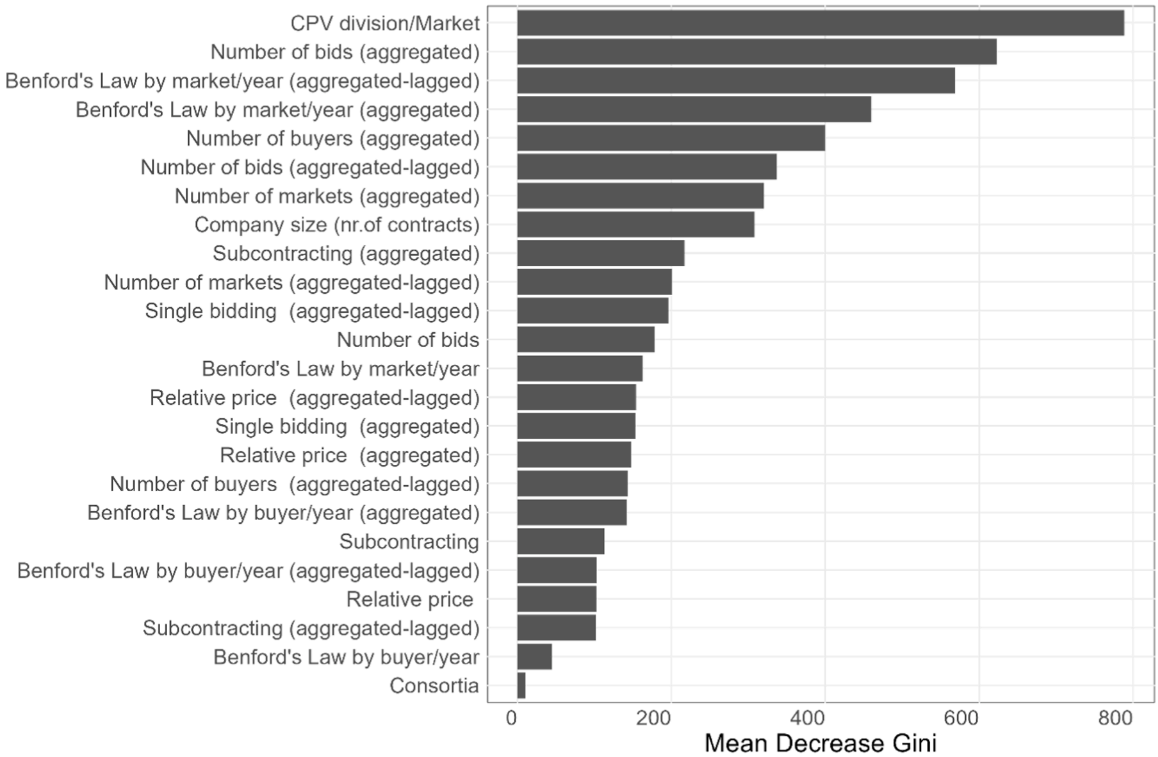


**Fig. 2.** Variable importance chart, best random forest model, all elementary collusion risk indicators plus company size and contract market, All countries, Full dataset
Notes: The importance plot ranks all model variables based on their contribution to decreasing the Gini impurity. It highlights the relative predictive power of each variable. Aggregate indicators—such as the number of bids, Benford's law compliance, number of buyers, and number of markets—are the most influential, followed by contract-level predictors.

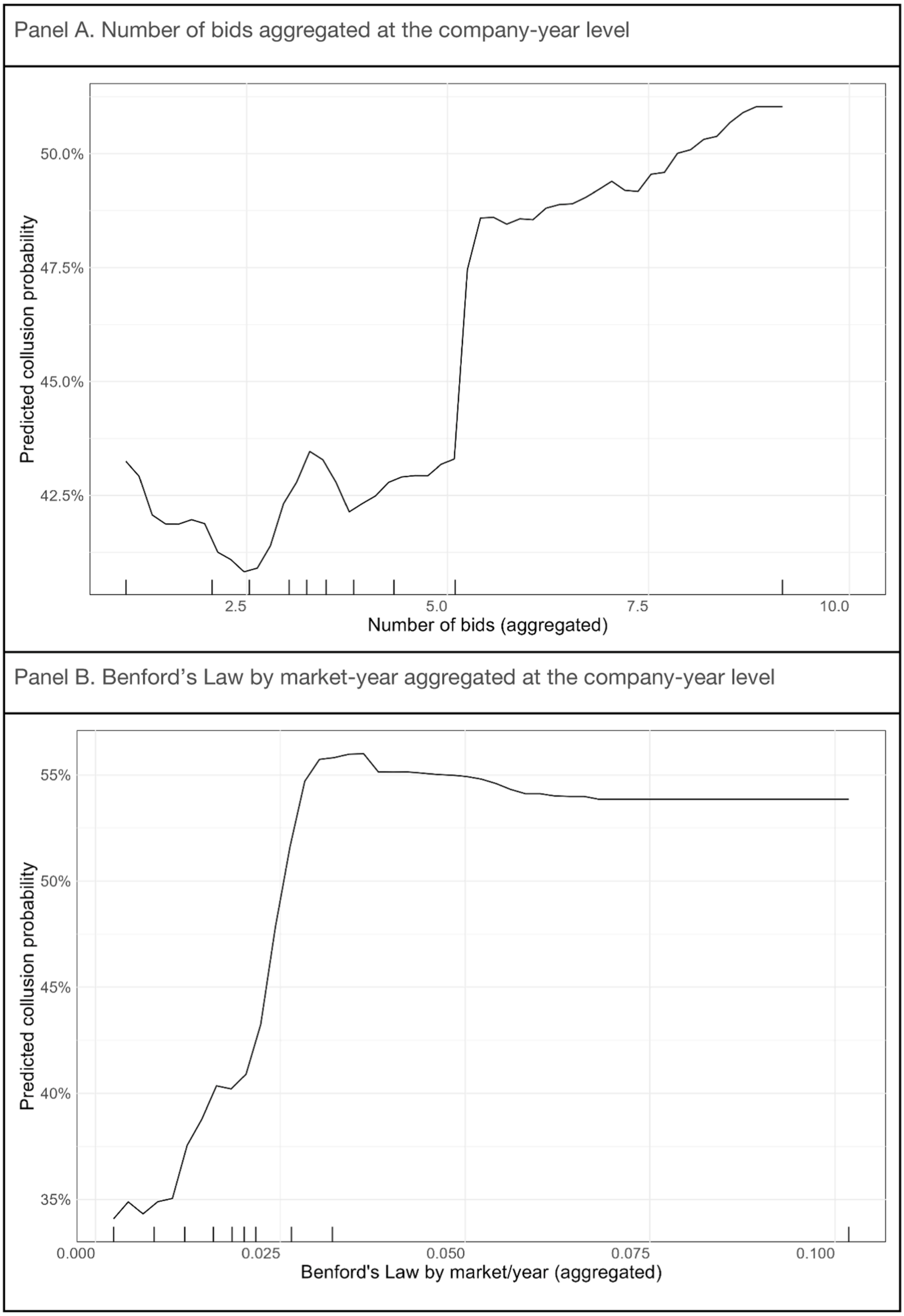


**Fig. 3.** Partial dependence plots for most important variables, best random forest model, all countries, training dataset of proven cases
Notes: Each panel shows the marginal effect of a key predictor on the predicted probability of collusion, holding other variables constant. Panel A: Number of bids (company-year); B: Benford's Law (market-year, company-year); C: Lagged number of buyers (company-year); D: Number of markets (company-year). Rug marks on the x-axis show the distribution of observed values.

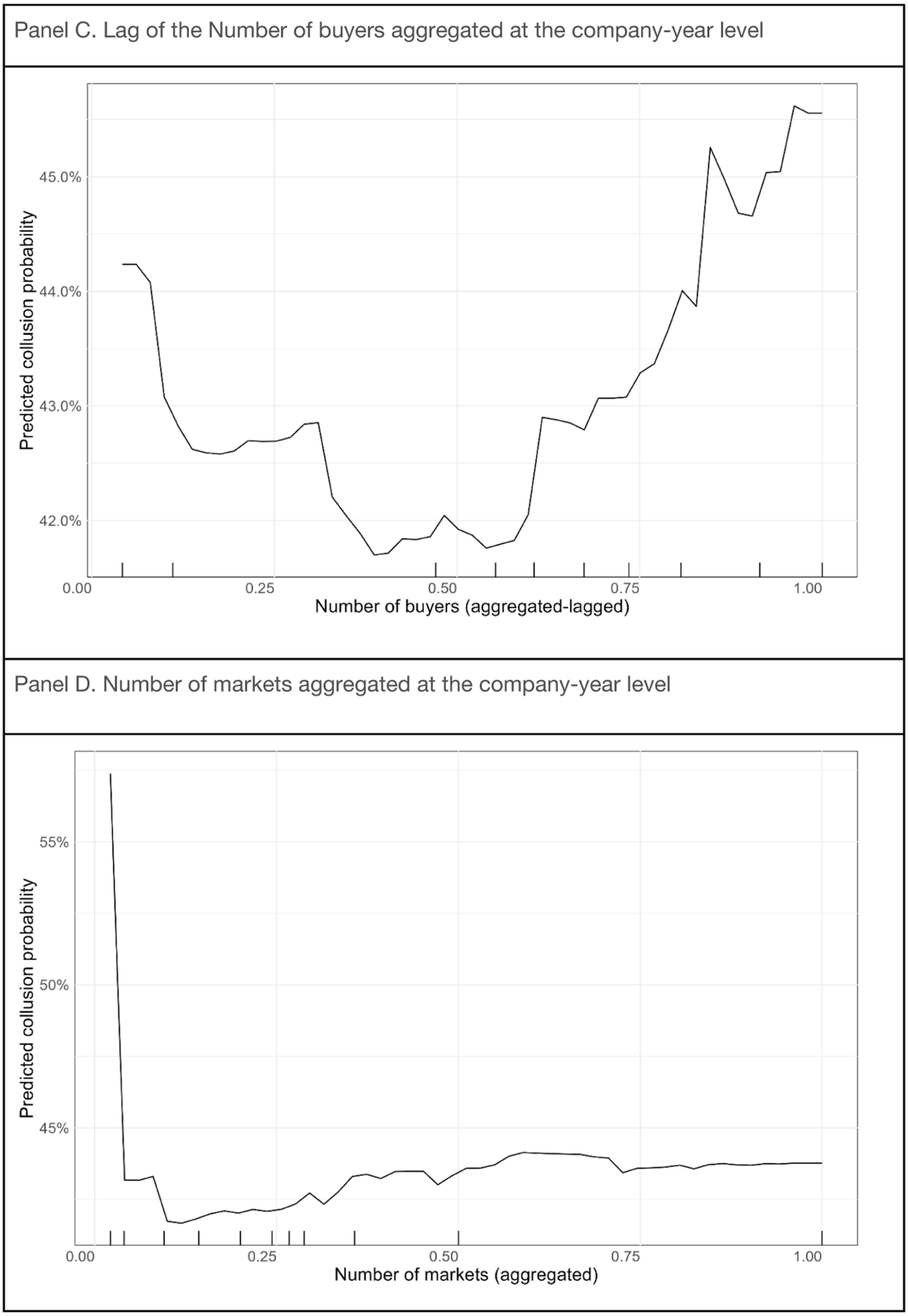


**Fig. 3.** (*continued*).

in year *t*), we see a U-shaped impact on predicted cartel risks with both very few bids (weaker effect) and very many bids (stronger effect) predicted to have high collusion risk (Fig. 3, Panel A).[26] Conceptually related indicators such as single bidding and number of bids (aggregated in *t-1*) show a similar pattern too (see Appendix E, Figure A.6, Panel A and B). The observed U-shaped relationship aligns with two distinct cartel strategies operating simultaneously in our data: either bid suppression (i.e., a low number of submitted bids) or the submission of an unusually high number of losing bids by cartel members to create the illusion of competition.

Second, regarding the Benford's Law compliance, aggregated by market and year in year *t*, we see a straightforward relationship in line with theory. The indicator measures the mean absolute deviation from Benford's distribution in market-year groups with larger deviations providing a stronger signal of collusion. The identified relationship is non-linear suggesting that low to medium values of mean absolute deviation are non-risky while a large uptick of about 20 percentage points in predicted cartel probability is estimated as the deviation crosses a threshold (Fig. 3, Panel B). A very similar relationship is identified for Benfords' Law in year *t-1* (See Appendix E, Figure A.6, Panel C).

Third, the number of distinct buyers from whom a supplier wins contracts from in year *t-1* (normalized by the number of contracts won) displays, again, a U-shaped relationship with predicted cartel risks (Fig. 3, Panel C) .[27] The model predicts a high cartel risk when a supplier secures contracts from a large number of buyers. In contrast, firms with a buyer count at or below the median exhibit lower cartel probabilities, but a notable uptick in cartel risk is observed at the lowest end of the scale. This suggests the presence of two simultaneous cartel strategies: either restricting bids to a select group of buyers (i.e., very few buyers per supplier-year) or strategically submitting numerous losing bids to a broad range of buyers.

Fourth, we look at an indicator conceptually similar to the previous one: the number of distinct markets in which a supplier participates in a year, normalized by the number of contracts won (Fig. 3, Panel D). Similar to some of the previous predictors, the model predicts higher cartel risks at both end of the distribution, although the effect is much more pronounced at the lower end. This pattern suggests that suppliers deviating significantly from the average number of markets tend to have a higher probability of collusive behaviour, especially when they are present on very few, typically only on one market. We include aggregated cartel screens with values in t and t-1 in order to allow the models to capture changes over time within the same company, we further explore the impacts of the number of distinct markets. Fig. 4 presents the SHAP (Shapley Additive Explanations)[28] scores for the number of distinct markets in t, with observations coloured based on the lagged (t-1) value of the same indicator. The pattern observed in the partial dependence plot (Fig. 3, Panel D) is confirmed and refined here. Companies which have consistently low numbers of distinct markets in multiple years have high cartel risks predicted. Interestingly, many companies which move from having presence in many markets (yellow and green dots) to participating in few markets only (i.e. indicating withheld bids) are also predicted to be of high risk - i.e. yellow and green dots above the 0 line and to the left of 0.5 on the y-axis.

## 5. Discussion

Having reviewed the main results of the different models and identified the best predictive model, we extrapolate risks for contracts with unknown cartel status and discuss policy implications.

### 5.1. Extrapolation

Taking the best Random Forest model identified above, we predict the risk of collusive behaviour in the full database consisting of around 3.3 million observed public procurement contracts in the seven countries during 2004–2021 (Fig. 5). Overall, the model predicts an average cartel risk score of 44 %. Approximately 21 % of all contracts observed across the seven countries receive a predicted cartel risk score below 20 %, while around 59 % of contracts fall below the 50 % risk threshold. Nevertheless, the upper third of contracts is predicted to have a cartel probability exceeding 64 %, suggesting a strong indication of collusive behaviour. This is roughly in line with findings in the literature using different data and indicators (e.g. Kawai and Nakabayashi, 2022). However, there is a strong cross-country variation in both the distribution of predicted cartel risks and the average risk per country (Fig. 6). Except for Spain, all countries are predicted to have only a minority of their public procurement markets likely collusive.

### 5.2. Policy implications

Enforcement authorities can improve public welfare with better detection and deterrence of cartels. Our methods have potential applications for both aims. For instance, improved cartel screening increases the risk of getting caught, and some cartels may decide to discontinue their collusion as a result. Similarly, individual whistleblowers may come forward if screens are thought to be improving competition authority effectiveness. Ongoing cartels may opt to weaken the behavioural traces they leave by decreasing their profit margins, a partial win for public welfare. Finally, improved screens increase coordination requirements; while in the 1950s colluding

[26] We adjusted the "number of bids" variable by treating outlier values as missing data to enhance the visibility of the pattern in the figures presented in Figure 3, Panel A, and Figure A.6, Panel B.

[27] Please note that we chose a less impactful variant of this variable as it better fits with theory, the more important variant aggregated in year t, can be found in Appendix E, Figure A.6, Panel D. It shows more of a positive, non-linear relationship than a U-shaped.

[28] SHAP (Shapley Additive Explanations) values provide a game-theoretic approach to interpreting machine learning models by quantifying the marginal contribution of each feature to an individual prediction.

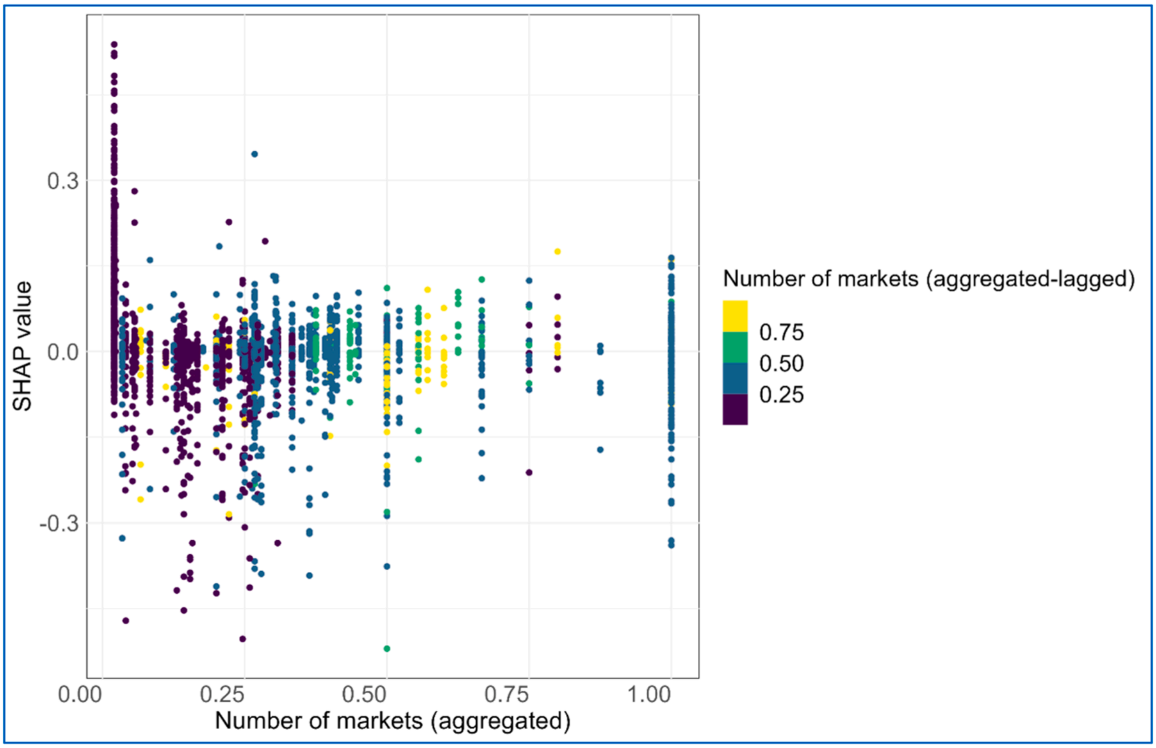


**Fig. 4.** SHAP interaction plot for the number of distinct markets indicators over time
Notes: SHAP values show how each feature contributes to the prediction. This plot illustrates the number of distinct markets a company operates in during year t, interacted with the same company's number of distinct markets in year t-1. Low market diversity signals high risk, while companies shifting from broad to narrow market presence (e.g. green/yellow points to the left) are also flagged as high risk—suggesting potential strategic bid withholding.

firms simply submit identical bids for asphalt contracts in Oklahoma (Funderburk, 1974), nowadays cartels employ complex strategies to evade detection and to keep all members satisfied. If the greater need for coordination increases communication among cartel members - to meet in person, or to otherwise reveal their behaviour - then they may become easier to catch. Together, these costs make collusion less attractive in the long run and may drive down the prevalence of cartels in general.

Based on the accuracy of our predictive models, comparable to the models commonly used in the literature, we see two policy implications: i) supporting investigation; and ii) informing preventive and pro-competition policy interventions. To improve investigations targeting public procurement cartels, we recommend the adoption of models like the ones developed in this article. We can conclude that both the theoretical and empirical advances in the literature have reached a level which warrants real life use. Applying such predictive models to regularly updated data, markets can be scanned to find high collusion risk markets. In addition, ongoing investigations could also be supported by predictive modelling where models could recommend further contracts of the investigated firms or further accomplice firms. Demonstrating this point, Fig. 7 presents scatter plots showing the relationship between the average probability of collusion and the logarithm of total spending (in billions)[29] across four countries: France, Hungary, Latvia, and Sweden. Each point represents a 3-digit CPV market with the red dashed lines showing the average values on each scale. Using such analysis, authorities can identify markets that are simultaneously high risk and high value, and consider investigating companies in those.

Authorities can deepen their investigations by analysing more detailed product classifications over time. Fig. 8 presents the annual probability of collusion for three country-market combinations, with each panel representing a distinct 4-digit CPV market within a specific country. Observing these trends over time reveals potential cartels forming or dissolving. For instance, Bulgaria's security systems market saw a peak in collusion probability around 2014, followed by a decline, which may indicate the dissolution of a cartel. Meanwhile, Hungary's maintenance services market exhibited a steady decline in collusion probability over time, suggesting a gradual reduction in cartel activity in that market.

As we find that about a third of the public procurement markets of seven European countries have high cartel risk, there is a strong argument for pro-competition, preventive policies. These policies could be tailored to address market entry and other barriers to competition especially in markets where collusion risks are found to be high.

[29] All contract prices are in EUR except for Hungarian contracts which are in Hungarian Forints.

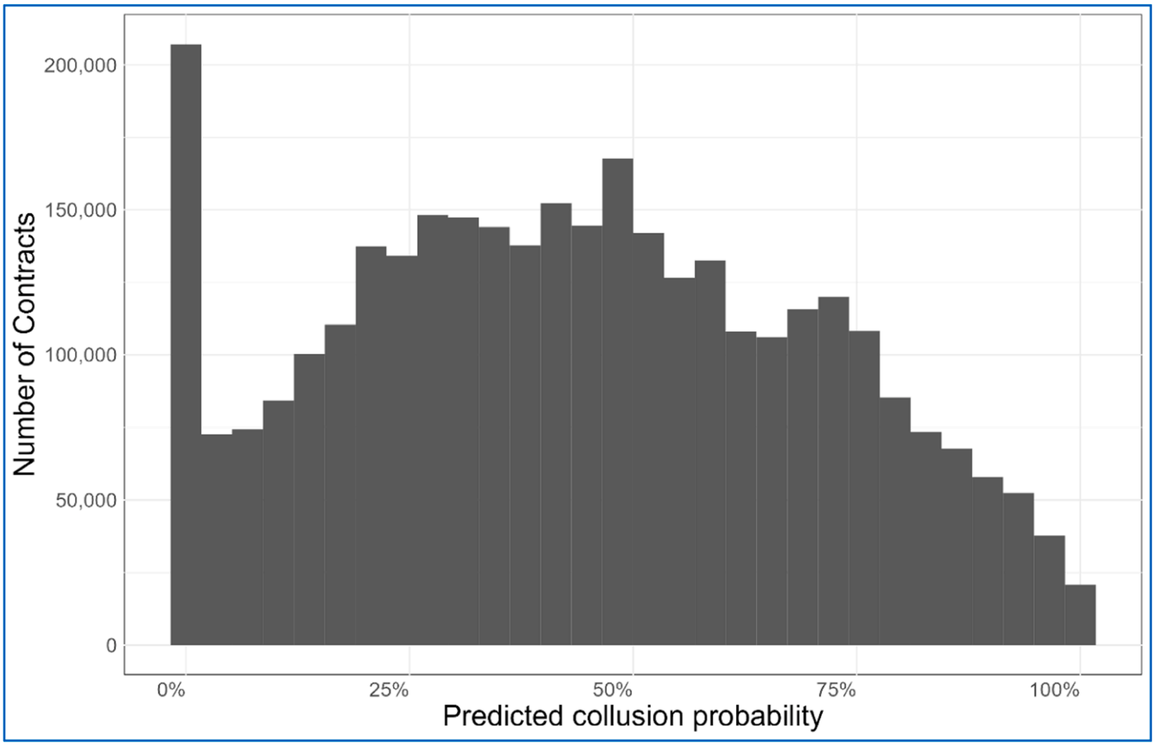


**Fig. 5.** Distribution of the predicted collusion probability, using the best random forest model, all countries, full dataset (including contracts with unknown cartel status).

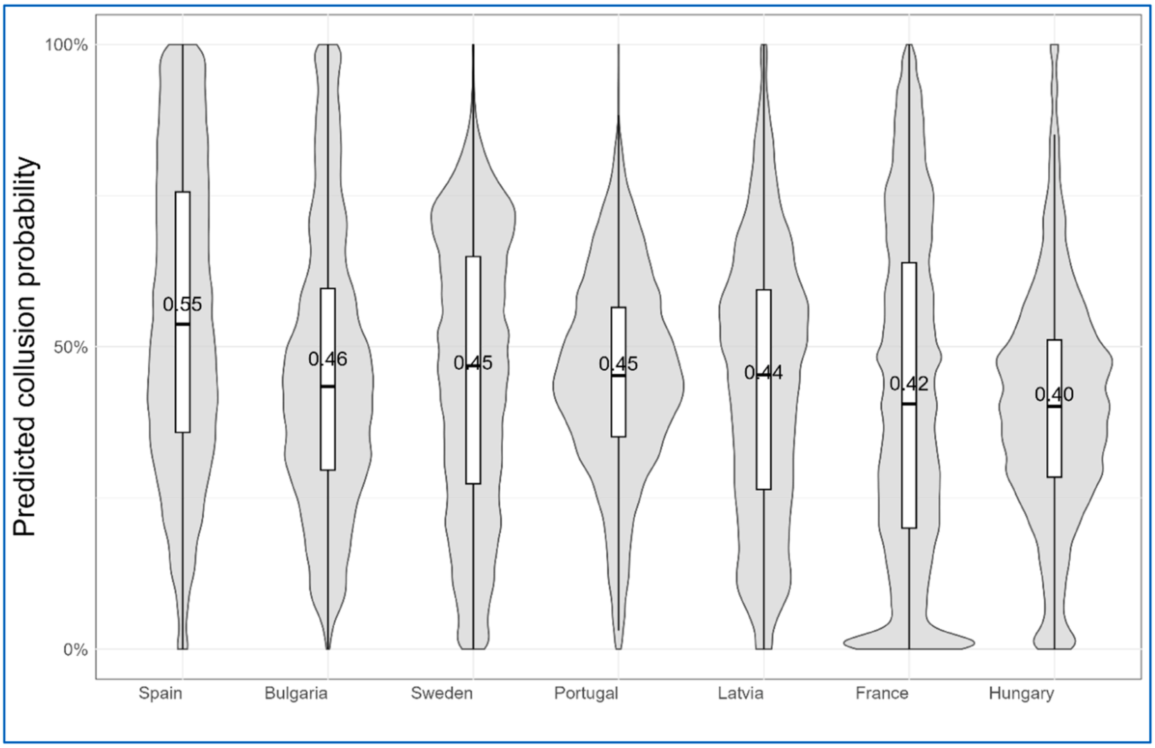


**Fig. 6.** Distribution of the predicted collusion probability by country, using the Best random forest model, All countries, full dataset (including contracts with unknown cartel status).

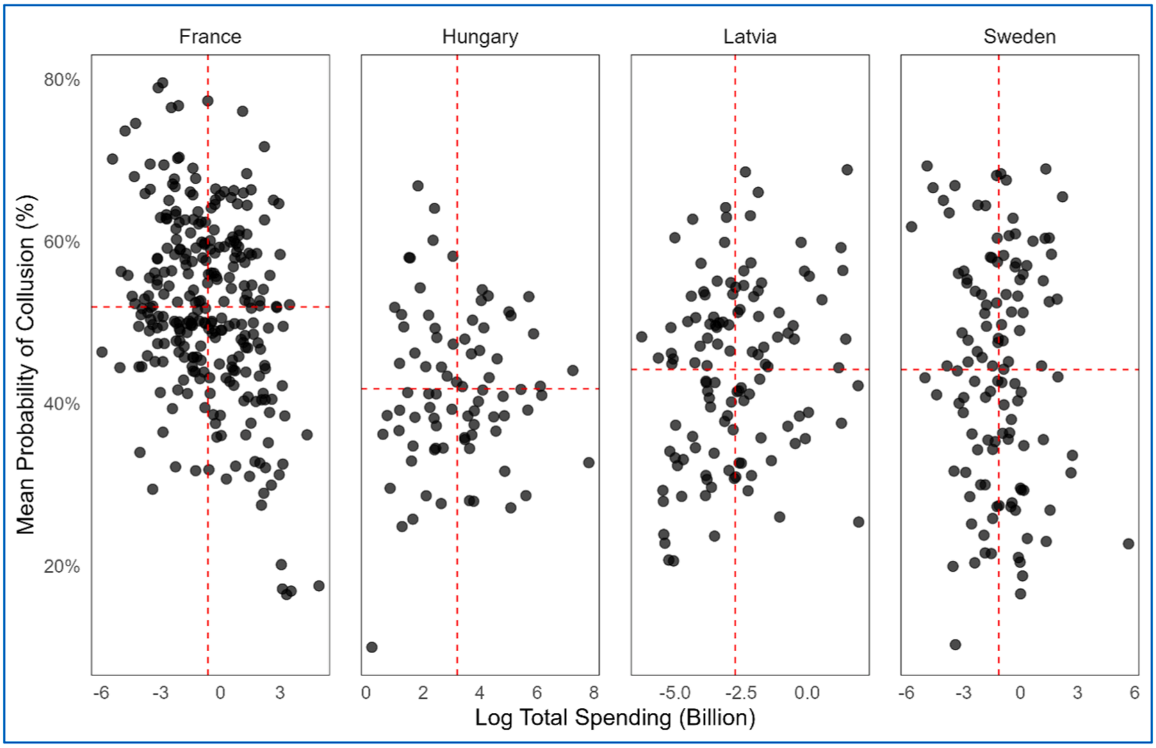


**Fig. 7.** Relationship between collusion probability and total spending by 3-digit cpv markets
Notes: Each point represents a 3-digit CPV market in France, Hungary, Latvia, or Sweden, plotting average predicted collusion probability against the log of total spending (in billions, local currency). Red dashed lines indicate country-level averages. This visualization demonstrates how flagging markets can help policymakers by identifying high-risk, high-value markets.

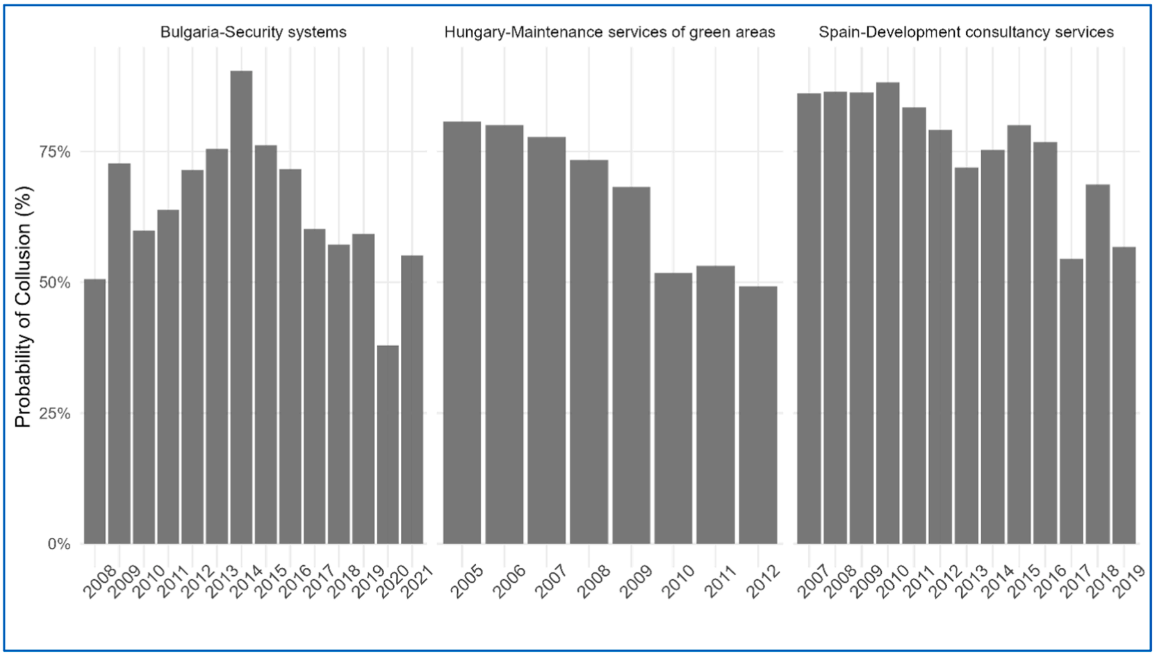


**Fig. 8.** Collusion probability trends over time in key 4-digit cpv markets.

The limitations of our modelling exercise have also revealed the need for improvements in public procurement data for better cartel risk modelling. There is a strong case for improving public procurement data quality and scope. In particular, the rate of missing data should be lowered in publicly available public procurement data repositories and government publication websites. Moreover, a range of key missing fields have prevented our models from incorporating additional collusion risk indicators. Among these, variables such as losing bidder and bid price information were missing in most countries. When it comes to data scope, several countries analysed only publicly disclosed higher value contracts which means that real-time monitoring of lower value contracts is very costly, if not impossible. Hence, recording all or nearly all public procurement transactions in centrally maintained e-procurement systems is of great value, not only for cartel risk detection but also for other uses of public procurement data (e.g. spending efficiency measurement).

There is a similarly strong case for building risk prediction and monitoring systems which continuously improve. That is, starting from the currently available best predictive models, competition authorities and other law enforcement bodies should feed the latest investigative results back into their learning models allowing them to adapt and improve. It is expected that cartels learn from past, especially recent, enforcement action, hence predictive models should reflect changes in cartel behaviours to stay relevant and accurate. The complexity of models and their continuous improvement would also make it harder for potentially collusive firms to strategically adapt and avoid detection. In addition, many of the adaptations needed by colluding firms are costly, for example lowering prices offered or more carefully that is more frequently coordinating bidding activities; hence a continuously improving monitoring framework lowers the incidence of collusion even if some strategic adaptation remains possible.

## 6. Conclusions

We set out to develop a high accuracy predictive model capable of tracking diverse cartel behaviours across many countries and over years. We drew on publicly available data for 73 cartels from seven countries during 2004–2021 and applied widely used machine learning methods, such as Random Forests, to combine diverse cartel screens into a theoretically sound model (find our replication package at: https://zenodo.org/records/17595875). Our best models achieve 70–84 % prediction accuracy across countries on unseen, test datasets of many proven cartels; while our prediction accuracy goes up to 93–97 % when tested on a single cartel case, typical of prior literature.

The high accuracy of the prediction when applied to held-out countries suggests our models learn generalizable patterns of collusion that go beyond idiosyncratic strategies particular to some countries or markets (see Appendix C.3 on the accuracy impact of adding new countries to the training set). Models are parsimonious, using only six elementary cartel risk indicators (although calculating their variants and different aggregates) and two control variables (company size and contract market). Unusually to many black-box machine learning approaches, our most important predictors are estimated to have an impact on predicted cartel probabilities in line with theory.

We also used the most accurate model to predict cartel probability for around 3.3 million contracts in seven European countries. This extrapolation suggests that about a third of public procurement contracts was awarded to firms engaging in potentially collusive behaviour. As our models were built on readily available, large-scale public procurement datasets, they can be applied to support competition enforcement, contributing to both investigations and pro-competition policy interventions.

Our approach nevertheless suffers from several limitations. First, the administrative data we could gather from publicly available sources, while large-scale, often has quality and scope issues. Many variables suffer from high missing rates and key variables suggested by the literature were not available at all (e.g. losing bid prices, company ownership). Second, any learning model can only learn from known cases which may represent a biased sample of the true range of cartel behaviours. For example, if more sophisticated cartels are harder to detect they may not show up in our learning dataset at all, so we under-estimate their presence in the data. Third, while tree-based classifiers handle noisy data well, they offer predictive rather than causal insights than for example fully specified econometric models. Certainly, further, more direct evidence is needed for a legal case. Fourth, model accuracy will drift as firms adapt to improved detection and data reporting formats change, requiring periodic retraining and threshold checks to keep the predictions accurate.

Future research could address the shortcomings of our approach. In particular, adding more countries, indicators (e.g. losing bid-based indicators) and cartel cases should improve on many limitations listed above. As more known cartel cases and countries are added to the model, we get closer to a genuinely generic cartel risk detection model which should be of great scientific and policy value.

## CRediT authorship contribution statement

**Mihály Fazekas:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Data curation. **Bence Tóth:** Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization. **Johannes Wachs:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Aly Abdou:** Writing – original draft, Validation, Software, Data curation.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijindorg.2025.103228.

## Data availability

https://zenodo.org/records/17595875 (The full replication package can be found at:)

## References


Abrantes-Metz, R.M., Froeb, L.M., Geweke, J., Taylor, C.T., 2006. A variance screen for collusion. Int. J. Ind. Organ. 24 (3), 467. –86.4.

Albano, G.L., Spagnolo, G., Zanza, M., 2009. Regulating joint bidding in public procurement. J. Compet. Law Econ. 5 (2), 335–360. http://jcle.oxfordjournals.org/content/5/2/335.short. May 13, 2016.

Alexander, B.J., 1997. Mechanisms for rent transfers: subcontracting among military aircraft manufacturers. Public Choice 91 (3), 251–269.

Asker, J., 2010. A study of the internal organization of a bidding cartel. Am. Econ. Rev. 100 (3), 724–762.

Athey, S., Bagwell, K., Sanchirico, C., 2004. Collusion and price rigidity. Rev. Econ. Stud. 71 (2), 317–349.

Bajari, P., Ye, L., 2003. Deciding between competition and collusion. Rev. Econ. Stat. 85 (4), 971–989.

Barrus, D.R., 2011. Firm Bidding Behavior in Highway Procurement Auctions: An Analysis of Single-Bid Contracts, Tacit Collusion, and the Financial Impact on Kentucky.

Bosio, Erica, Djankov, Simeon, Glaeser, Edward, Shleifer, Andrei, 2022. Public procurement in law and practice. Am. Econ. Rev. 112 (4), 1091–1117.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Carbone, C., Calderoni, F., Jofre, M., 2024. Bid-rigging in public procurement: cartel strategies and bidding patterns. Crime Law Soc. Change avail. online.

Chassang, S., Ortner, J., 2019. Collusion in auctions with constrained bids: theory and evidence from public procurement. J. Polit. Econ. 127 (5), 2269–2300.

Conley, T.G., Decarolis, F., 2016. Detecting bidders groups in collusive auctions. Am. Econ. J.: Microecon. 8 (2), 1–38.

Estache, A., Iimi, A., 2008. Procurement Efficiency for Infrastructure Development and Financial Needs Reassessed, 4662. World Bank policy research working paper.

Fazekas, M., Tóth, B., 2016. Assessing the Potential for Detecting Collusion in Swedish public Procurement. Swedish Competition Authority, Stockholm.

Fazekas, M., Cingolani, L., Tóth, B., 2018. Innovations in objectively measuring corruption in public procurement. Governance indicators. Approaches Prog. Promise 154–185.

Fazekas, Mihály, Tóth, Bence, Wachs, Johannes, 2023. Public Procurement cartels: A large-Sample Testing of Screens Using Machine Learning. Government Transparency Institute, Budapest. GTI-WP/202302.

Fazekas, Mihály, Bence, Tóth, Abdou, Aly, Al-Shaibani, Ahmed, 2024. Global contract-level public procurement dataset. Data Br. 54.

Funderburk, D.R., 1974. Price fixing in the liquid-asphalt industry: economic analysis versus the hot document. Antitrust L. Econ. Rev. 7, 61.

Hewitt, C., McClave, J., Sibley, D., 1996. Incumbency and Bidding Behavior in the Dallas-Ft. worth school milk market. Unpublished manuscript.

Huber, M., Imhof, D., 2019. Machine learning with screens for detecting bid-rigging cartels. Int. J. Ind. Organ. 65, 277–301.

Huber, M., Imhof, D., Ishii, R., 2022. Transnational machine learning with screens for flagging bid-rigging cartels. J. R. Stat. Soc.: (Stat. Soc.) 1074–1114.

Ishii, R. (2008). Collusion in repeated procurement auction: a study of a paving market in Japan. ISER Discussion Paper No. 710, https://doi.org/10.2139/ssrn.1148064.

Ishii, R., 2009. Favor exchange in collusion: empirical study of repeated procurement auctions in Japan. Int. J. Ind. Organ. 27 (2), 137–144.

James, G., Witten, D., Hastie, T., Tibshiranim, R., 2021. An Introduction to Statistical Learning: With Applications in R, 2nd edition. Springer.

Kawai, K., Nakabayashi, J., Ortner, J.M., Chassang, S., 2022. Using bid rotation and incumbency to detect collusion: a regression discontinuity approach. NBER Working Papers 29625. National Bureau of Economic Research.

Kawai, K., Nakabayashi, J., 2022. Detecting large-scale collusion in procurement auctions. J. Polit. Econ. 130 (5), 1364–1411.

Lanzillotti, R.F., 1996. The great school milk conspiracies of the 1980s. Rev. Ind. Organ. 11 (4), 413–458.

Levenstein, M.C., Suslow, V.Y., 2006. What determines cartel success? J. Econ. Lit. 44 (1), 43–95.

Lyra, M.S., Damásio, B., Pinheiro, F.L., Bacao, F., 2022. Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. Appl. Netw. Sci. 7 (1), 83.

Mena-Labarthe, C., 2012. Mexican experience in screens for bid-rigging. Antitrust Chron. 3.

Morozov, I., Podkolzina, E., 2013. Collusion Detection in Procurement Auctions. National Research University Higher School of Economics. No. WP BRP 25/EC/2013.

Morselli, C., Ouellet, M., 2018. Network similarity and collusion. Soc Netw. 55, 21–30.

OECD, 2014. Detecting Bid Rigging in Public Procurement. OECD, Paris. See. http://www.oecd.org/competition/cartels/42594486.pdf.

Oxera, 2013. Hide and Seek: The Effective Use of Cartel Screens. Downloaded: 2021.01.19h.t.tp://w.w.w.oxera.com/getmedia/210bc5bc-0cc9-40ea-8bc9-6c8b2406b485/Cartelscreens.pdf.aspx?ext=.pdf.

Pesendorfer, M., 2000. A study of collusion in first-price auctions. Rev. Econ. Stud. 67 (3), 381–411.

Porter, R.H., Zona, J.D., 1999. Ohio school milk markets: an analysis of bidding. Rand J. Econ. 263–288.

Rodríguez, M.J.G., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P.E.D., Signor, R., 2022. Collusion detection in public procurement auctions with machine learning algorithms. Autom. Constr. 133, 104047.

Samà, D.. Cartel detection and collusion screening: an empirical analysis of the London Metal Exchange. MPRA Paper No. 55363. https://mpra.ub.uni-muenchen.de/55363/.

SCA, 2015. Honesty Pays: How to Detect and Prevent Bid-Rigging Cartels.

Schwalbe, U., 2018. Algorithms, machine learning, and collusion. J. Compet. Law Econ. 14 (4), 568–607.

Scott, F.A., 2000. Great school milk conspiracies revisited. Rev. Ind. Organ. 17 (3), 325–341.

Tóth, B., Fazekas, M., Czibik, Á., Tóth, I.J., 2014. Toolkit for Detecting Collusive Bidding in Public procurement: With examples from Hungary. Government Transparency Institute, Budapest. GTI-WP/2014:02.

Vadász, P., Benczúr, A., Füzesi, G., Munk, S., 2016. Identifying illegal cartel activities from open sources. Open Source Intelligence Investigation. Springer, Cham, pp. 251–273.

Wachs, J., Kertész, J., 2019. A network approach to cartel detection in public auction markets. Sci. Rep. 9 (1), 1–10.

World Bank, 2009. Fraud and Corruption. Awareness Handbook. World Bank, Washington DC.

World Bank, 2011. Curbing Fraud, Corruption, and Collusion in the Roads Sector. World Bank, Washington DC.