Computational Economics
https://doi.org/10.1007/s10614-022-10315-w

A Machine Learning Approach for Flagging Incomplete Bid-Rigging Cartels

Hannes Wallimann[1,2] · **David Imhof**[2,3,4] · **Martin Huber**[2]

Accepted: 19 August 2022
© The Author(s) 2022

Abstract
We propose a detection method for flagging bid-rigging cartels, particularly useful when cartels are incomplete. Our approach combines screens, i.e., statistics derived from the distribution of bids in a tender, with machine learning to predict the probability of collusion. As a methodological innovation, we calculate such screens for all possible subgroups of three or four bids within a tender and use summary statistics like the mean, median, maximum, and minimum of each screen as predictors in the machine learning algorithm. This approach tackles the issue that competitive bids in incomplete cartels distort the statistical signals produced by bid rigging and it outperforms previously suggested methods in applications to incomplete cartels based on empirical data from Switzerland.

Keywords Bid rigging detection · Screening methods · Descriptive statistics · Machine learning · Random forest · Lasso · Ensemble methods

JEL Classification C21 · C45 · C52 · D22 · D40 · K40 · L40 · L41

✉ David Imhof
david.imhof@weko.admin.ch

Hannes Wallimann
hannes.wallimann@hslu.ch

Martin Huber
martin.huber@unifr.ch

[1] Competence Center for Mobility, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland

[2] Department of Economics, University of Fribourg, Fribourg, Switzerland

[3] Swiss Competition Commission, Hallwylstrasse 4, 3003 Bern, Switzerland

[4] UniDistance, Faculté d'économie, Brig, Switzerland

Published online: 11 September 2022

1 Introduction

When firms deviate from competitive behavior and instigate a cartel, they secretly conspire to raise prices or lower the quality of goods or services. As such, conspiracies directly harm taxpayers, buyers, or sellers. Cartel formation remains a pervasive problem and has been considered in a range of studies. See for instance the Swedish asphalt cartel described in Bergman et al. (2020), collusion among seafood processors in the US (Abrantes-Metz et al. 2006), bid rigging in public procurement auctions for construction works in Japan (Ishii 2014), in Poland (Foremny et al. 2018), in Canada (Clark et al. 2018) and in the US (Porter and Zona 1993; Feinstein et al. 1985) and bid rigging for school milk contracts in Ohio (Porter and Zona 1999), Florida and Texas (Pesendorfer 2000). To enhance the fight against cartels, the OECD recommends competition agencies to promote pro-active methods for uncovering conspiracies, as such methods may help to discover cartels where leniency is unlikely to be sought (OECD 2013). Answering the need for statistical tools in this context, Porter and Zona (1993), Bajari and Ye (2003), Harrington (2008), Jimenez and Perdiguero (2012), Imhof et al. (2018), Crede (2019) and Bergman et al. (2020), among others, have proposed different methods for uncovering cartels.

However, the detection of cartels might be more challenging in the presence of competitive bidders participating in markets in which a cartel is active (McAfee and McMillan 1992; Hendricks et al. 2008; Asker 2010; Bos and Harrington 2010; Conley and Decarolis 2016; Decarolis et al. 2020). When a cartel is incomplete due to competitive bidders, it weakens the statistical pattern produced by bid rigging in the distribution of bids, increasing the difficulty of detecting a cartel. Moreover, a cartel might temporarily collapse because of deserters, i.e., it is not always stable. This instability in the cartel might affect the screens rendering the statistical signals of bid rigging more challenging to detect. Finally, a cartel aware of methods for uncovering cartels might try to weaken the statistical pattern due to bid rigging in order to decrease the ability of such methods to predict the cartel's presence.

Thus, this paper offers an original application of a detection method based on screens to detect both incomplete and complete bid-rigging cartels. Screens are statistics derived from the distribution of bids in a tender capturing the distributional changes produced by bid rigging (see Abrantes-Metz et al. 2006; Hueschelrath and Veith 2014; Abrantes-Metz et al. 2012; Jimenez and Perdiguero 2012; Imhof et al. 2018; Imhof 2019). Our novel approach consists of calculating screens for all possible subgroups of three or four bids in a tender and not only for all bids. We then use the screens calculated for all the subgroups in a particular tender to calculate descriptive statistics of each screen, which synthesize the properties of the distribution of bids in a tender. Those descriptive statistics of screens, henceforth called 'summary screens', circumvent the distortion that competitive bidders or deserters generate in the statistical signals produced by bid rigging in a tender, rendering our suggested method robust to the presence of competitive bidders.

In our study, we combine the summary screens with machine learning as in a prediction policy problem (see Kleinberg et al. 2015). Machine learning has been

applied in a rapidly increasing number of studies (Rabuzin and Modrusan 2019; Imhof and Wallimann 2021; García Rodríguez et al. 2020; Rodríguez et al. 2022; Silveira et al. 2022; Huber et al. 2022) and aims at finding the optimal combination of covariates that best predicts the presence or absence of bid rigging in a tender. Also related to our paper is the recent study of Uslu et al. (2021), applying machine learning to investigate trade-based manipulations of capital market instruments. Moreover, our paper is related to studies analyzing bidding strategies (see, e.g., Liu et al. 2020; Cai et al. 2019) and applying predictive models (see, e.g., Mir et al. 2020; Mirzapour et al. 2019) in other research fields. As we focus on the predictive performance, we do not have to construct explicit structural models for collusion. To train and evaluate models, we focus on the random forest (see Breiman 2001) as machine learner because it provides a flexible prediction method that does not impose any parametric (e.g., linearity) assumptions when considering our large set of screens. In contrast to many other machine learners, random forests do not require tuning specific penalty terms, see the discussion in Athey and Imbens (2019), and are therefore easier to be implemented. This appears desirable if a competition agency applies our detection method for screening procurement markets.

Calculating screens for subgroups as in our approach is also considered in Conley and Decarolis (2016) and Chassang et al. (2022). First, Conley and Decarolis (2016) investigate subgroups to detect cartels in collusive auctions in Italy, but in contrast to our method (which considers all possible subgroups in a tender), they exploit firm-specific covariates (such as, e.g., common owner, municipality, or country) to form subgroups. Relying on firm-specific covariates could impede a broad screening activity if firm-specific data are unavailable or if the time needed to collect them without raising the attention of potential cartel participants is lacking. Chassang et al. (2022) show that winning bids tend to be isolated in terms of value when bidders collude. For analyzing the missing density of bids between the first and the second-lowest bids, they calculate the normalized margin. First, bids are normalized by the reserve price (which would be impossible in our data). Second, they calculate for each normalized bid i the difference with the minimal normalized bid (other than i) in each tender. A missing density around zero for the normalized margin is incompatible with competition, especially if repeatedly observed. A competitive bidder rationally maximizing profits would be tempted to increase her or his bids if she or he regularly observes that other bidders submit substantially higher bids. Therefore, a density gap around zero is incompatible with competition.

Two important arguments favor of our approach based on machine learning and synthetized screens. First, it exclusively relies on information about bids rather than firm-specific characteristics or cost-related variables required for econometric tests (see for instance Bajari and Ye 2003; Aryal and Gabrielli 2013). Our suggested method requires only bid summaries, which are either public or readily accessible for competition agencies and thus not as costly to acquire as firm- or cost-specific information. The necessity to gather firm-level information can attract, in some cases, the attention of the cartel, decreasing the chance of success in acting against it. Second, machine learning relies on the hypothesis that bid rigging affects the distribution of bids in a tender (also common to other methods for flagging bid-rigging cartels as the econometric tests suggested by Bajari and Ye (2003)) but

remains agnostic about how the distribution is affected. In our case, it is sufficient to assume that bid rigging modifies the distribution of bids and that screens can capture these changes.

Our study investigates the correct classification rates of different methods in the context of incomplete cartels. We first apply a benchmark method, suggested by Imhof et al. (2018), which implements two screens with benchmarks, i.e., a rule of thumb, for classifying a tender as collusive or competitive. The second method applies machine learning using a set of screens, calculated based on all bids in a tender, so-called 'tender-based screens', to predict collusion. Finally, the third method is the novel approach suggested in this paper, which includes summary statistics of the screens (median, mean, maximum and minimum) calculated for all possible subgroups of bids in a tender as predictors in the random forest.

We use data from Switzerland, where the incidence of collusive and competitive tenders is known. We apply our approach to two investigations of the Swiss competition commission (hereafter COMCO): See-Gaster and Strassenbau Graubünden. Both cases were characterized by well-organized bid-rigging cartels, which sometimes faced competition from outsiders. These competitive bidders might have tried on one hand to benefit from the umbrella effect of the cartel by bidding higher than they would have done in a competitive situation (Bos and Harrington 2010). On the other hand, too many competitive bidders could have destabilized the formation of cartels.

We find that the benchmarking approach exhibits low correct classification rates for incomplete cartels. Using tender-based screens in predictive models, we obtain correct classification rates from 61 and 77% when competitive bidders are present. Applying our novel approach based on summary screens increases the performance to correct classification rates ranging between 67 and 84%. Further, we note that the performance of machine learning decreases with the proportion of competitive bids. This result confirms the findings from the investigations that cartel participants partially endogenize the presence of competitive bidders by adopting a more competitive behavior, at least in some cases.

The remainder of this study is organized as follows. Section 2 presents the bid-rigging cartels uncovered in Switzerland from which our data are drawn. Section 3 outlines the detection methods for flagging both complete and incomplete bid-rigging cartels. Section 4 applies our original application to incomplete cartels based on empirical data from the cases of See-Gaster and Strassenbau Graubünden. Section 5 concludes.

2 Bid-Rigging Cartels and Data

The Swiss Parliament revised the federal Cartel Act and introduced a sanction regime in April 2004, with an adaptation period of one year, alongside a compliance program. This legislative modification helped initiating a change in the praxis towards economically harmful bid-rigging cartels. At the end of 2004, COMCO began investigating the Ticino cartel, releasing its decision in 2007. The Ticino cartel dissolved without sanctions since it had ended its illegal conduct precisely before

Table 1 Decisions of COMCO in bid-rigging cases

Decisions of COMCO (excerpt)	Year	Sanctions (million CHF)	Participants
Road asphalting in Ticino	2007	–	17
Electric installations Bern	2009	1.2	7
Road construction and civil engineering Aargau	2011	7	18
Road construction and civil engineering Zurich	2013	0.5	12
Tunnel cleaning	2015	0.16	3
Road construction and civil engineering see-Gaster	2016	5	8
Construction in Val Mustair	2017	–	5
Six short decisions in Engadine	2017	1	12
Construction in lower Engadine	2018	7.5	7
One short decision in Engadine	2019	0.5	3
Road construction Graubünden	2019	11	12

April 2005, consuming the entire adaptation period. However, it stressed the damage and mischief of a bid-rigging cartel with a price increase of over 30% (see Imhof 2019). In 2008, COMCO decided to prioritize fighting bid rigging.

Following its decision in the Ticino case, the authority prosecuted many bid-rigging cases. Initially, COMCO rendered an essential decision against bid rigging every other year. From 2015 onwards, however, COMCO rendered more decisions, emphasizing its determination to prosecute bid-rigging conspiracies. Table 1 lists COMCO's most important decisions in bid-rigging cases and the sanctions it imposed in each case.

Overall, COMCO opens an investigation if there are reasonable grounds to assume the existence of a bid-rigging cartel. Compliance programs, whistleblowers, and procurement agencies can provide insightful information leading to the opening of an investigation. However, COMCO decided to reduce its dependence on such sources and started to develop statistical methods for detecting bid rigging based on screens (see also Imhof et al. 2018). Based on the latter method, COMCO opened an investigation of bid rigging in the region of See-Gaster in 2013.

Considering the evolution of the cases investigated by the COMCO in recent years, incomplete bid-rigging cartels occur more often than well-organized complete cartels. Therefore, if COMCO desires to reduce the dependence on external sources to open investigations, it must continue to improve its detection methods. Our approach for flagging both incomplete and complete bid-rigging cartels proposed in this paper responds to that need. It is likely to be of interest to competition agencies around the world.[1]

In the empirical analyses, we use data from two of Switzerland's most important cases: the See-Gaster cartel and the Graubünden asphalt cartel. Finally, after

[1] Another recent example is the new Procurement Collusion Strike Force (PCSF) in the US, which focuses on deterring, detecting, investigating, and prosecuting antitrust crimes, such as bid-rigging conspiracies and related fraudulent schemes. See https://www.justice.gov/opa/pr/justice-department-announces-procurement-collusion-strike-force-coordinated-national-response (accessed on June 6, 2021).

discussing procurement in Switzerland, we synthesize the main aspects of Swiss procurement data in both cases.

2.1 Procurement Data

Procurement agencies of cantons and cities in Switzerland follow the Agreement on Public Markets, which states that procurement agency can choose between four procedures: the open, the invitation, the selection, and the discretionary procedure.[2] In the construction sector, a procurement agency generally uses either the open procedure or the procedure on invitation. The open procedure does not restrict the participation of submitting firms, in contrast to the procedure on invitation, where the procurement agency invites only a small number of firms, in general, three to five, to submit a bid. This changes the nature of the competition, as the participating firms are aware of the restricted number of potential competitors.

A procurement agency announces future contracts and the deadline for submitting bids (varying according to the procedure) in an official journal. If a firm is interested in submitting, the procurement agency provides the firm with all the relevant documents or information for the contract. Firms prepare their bids for submission between the time of the announcement and the deadline. Collusive agreements, if any, between firms are typically concluded during this period.

At a pre-announced date, the procurement agency gathers the incoming bids for the contract and opens them. It officially records all the bids received on time in a bid summary or so-called official record of the bid opening and registers the firms' names, addresses, and bids. Having registered the official record of the bid opening, the procurement agency proceeds with a detailed examination of the bids. In awarding the contract, the agency considers not only the price of the bids, but also other criteria such as quality, references and environmental or social aspects. However, as contracts are relatively homogeneous in the construction sector, especially in road construction and associated civil engineering, the price in practice remains the most important criterion for awarding the contract. Furthermore, the differences in firms' criteria other than price are typically small. We, therefore, consider the procurement process as an almost first-price sealed-bid auction.

2.2 The Cartel in See-Gaster

COMCO opened its investigation in the region of See-Gaster mainly because of a statistical analysis based on procurement data from 2004 to 2010 provided by the canton of St. Gallen (see Imhof et al. 2018).[3] In total, eight firms participated in bid-

[2] The selection procedure allows the procurement agency to select and qualify a set of bidders to participate in a tender. This procedure is useful when bidders are too numerous, for example, in architectural design, where hundreds of architects are interested in bidding for the project. However, such a high number of bidders is rarely a problem in the construction sector.

[3] Report release: see decision *Bauleistung See-Gaster: Verfügung vom 8. Juli 2016*, available on the following internet page: https://www.weko.admin.ch/weko/fr/home/actualites/dernieres-decisions.html.

rigging conspiracies in the region of See-Gaster, including the district of See-Gaster in the canton of St. Gallen and the districts of March and Höfe in the canton of Schwyz.[4] Cartel participants regularly met once or twice a month. They discussed future contracts being put out to tender in their meetings and exchanged their interest in them. The contracts included road construction, asphalting and civil engineering. Before each meeting, one cartel participant sent an actualized table to all the others, listing all future contracts in the region of See-Gaster. Each cartel participant had a column to put a star to a contract if it was interested in obtaining the contract, or two stars if it wished to register a very high interest.[5] When the tender procedure for a contract started, the cartel typically designated the cartel participant that should win it. The allocation mechanism was based on the interests that had been announced and fairness in allocating contracts to participants for maintaining cartel stability.[6] In addition, if two cartel participants had both put two stars for a specific contract, they might have formed a consortium to share the contract, while other participants covered the consortium.[7]

The cartel took decisions on contract allocation during the meetings in which they discussed the list, but they organized separate meetings to discuss the price of the bids.[8] One reason for separate meetings is that not all cartel participants were interested in fixing the price since not all necessarily participated in the tender. Second, discussions about price might have taken up too much time, such that the cartel preferred the designated winner to invite the other bidders to a separate meeting for discussing the price. COMCO found some evidence that from time to time, the cartel used the mechanism of the mean in determining the bid to be made by the designated winner,[9] which implies that the latter had to submit either its own bid or the mean of all the exchanged bids in the separate meetings. Using this mechanism, the designated winner had some incentive to provide a relatively high bid to influence the calculated mean in the separate meeting. All the other cartel participants whose announced bids were below the mean or below the winner's bid increased their bids to cover the designated winner. As a result, they generally ensured a minimal price difference of 2–3% between the bid of the designated winner and their own bids.[10]

Finally, the cartel also made decisions about contracts that were left free for competitive bidding.[11] This decision was also determined by the presence of external bidders. When the extent of external bidders increased, the chances of a cartel's success decreased, and, thus, the incentive to collude declined. This was the case for

[4] See decision *Bauleistung See-Gaster: Verfügung vom 8. Juli 2016*, available on the following internet page: https://www.weko.admin.ch/weko/fr/home/actualites/dernieres-decisions.html.

[5] See decision *Bauleistung See-Gaster*, R. 809 ff.

[6] See decision *Bauleistung See-Gaster*, R 587, R 608 and R 623.

[7] See decision *Bauleistung See-Gaster*, R 620 ff. and R. 645.

[8] See decision *Bauleistung See-Gaster*, R. 649 ff.

[9] See decision *Bauleistung See-Gaster*, R. 714 ff.

[10] See decision *Bauleistung See-Gaster*, R. 714 ff. and R 718.

[11] See decision *Bauleistung See-Gaster*, R 681 ff. and R. 815 ff.

some high value contracts, for which more non-cartel firms were interested in bidding. Sometimes, the cartel also tried to bring external firms into the agreement.

In June 2009, the cartel ended its illegal conduct after COMCO launched house searches in the canton of Aargau, which to a certain extent explained the breakdown of the cartel. In its decision, COMCO attested that the cartel had discussed more than 400 contracts in the region of See-Gaster from 2004 to 2009 with a value of 198 million CHF. COMCO also proved that the cartel had attempted to rig at least 200 contracts with a value of 67.5 million CHF.[12] In making its decision, COMCO sanctioned the firms involved for bid-rigging conspiracies with more than five million CHF. Two firms applied to the leniency program, and two other firms settled an agreement to close the case. Four firms appealed against the decision.

2.3 The Strassenbau Cartel in Graubünden

The local trade association members organized the cartel in the canton of Graubünden for road construction. In its decision, COMCO proved that the cartel participants met regularly in the period being investigated, from 2004 to the end of May 2010. The meetings, called "allocation meetings" or "calculation meetings", were mainly held at the beginning of the year since the canton and the local municipalities put most of their contracts out to tender in the spring of each year.[13] The cartel discussed contracts for road construction and asphalting tendered by the canton of Graubünden and the local municipalities. Since mountains and valleys profoundly mark the geography of Graubünden, the cartel was divided into firms operating in the north and south, respectively.

In the north of Graubünden, the cartel mostly organized its meetings in the office of the most important mixing plant in the canton and, to a lesser extent, in the offices of the cartel participants. The meetings included either all of the twelve to thirteen cartel participants[14] or two different subgroups.[15] In the south, the total of six cartel participants[16] also organized such meetings, though changing their locations.

COMCO stated in its press release that the cartel decided upon the allocation of contracts based on a contingent determination for all the cartel participants in the canton of Graubünden.[17] The cartel allocated contracts according to the interests of each firm and fixed the price of the designated winner following a specific calculation method.[18] The price of the designated winner was usually above the minimal bid announced in the respective meeting. The calculation method, therefore, contributed to raising the price.

[12] See decision *Bauleistung See-Gaster*, R. 797 ff. and table 15.

[13] See decision *Strassenbau Graubünden*, R. 139.

[14] See decision *Strassenbau Graubünden*, R 247 ff.

[15] See decision *Strassenbau Graubünden*, R 195 ff.

[16] See decision *Strassenbau Graubünden*, R 248.

[17] See press release: https://www.newsd.admin.ch/newsd/message/attachments/58229.pdf.

[18] The online published decision *Strassenbau Graubünden* and the press release currently give no details of the calculation method.

During the period investigated, from 2004 to the end of May 2010, the cartel distributed 70% to 80% of the total value of the cantonal and communal road construction contracts. The cartel rigged approximately 650 road construction contracts concerning with a total value of 190 million CHF of market volume.[19] The cartel ceased its illegal conduct in the summer of 2010 since some firms decided to stop, mainly because of increasing concerns regarding the Cartel Act.[20]

2.4 Data from the Cases See-Gaster and Graubünden

We requested data on all bid summaries from the investigations of See-Gaster and Graubünden based on the Federal Act on Freedom of Information in the Administration (Freedom of information Act, FoIA).[21] COMCO approved the request and sent us the data, referred to hereafter as the Swiss data. They contain the bids, a running number for each contract, a dummy variable for each of the anonymized cartel participants and a dummy variable indicating whether the tender took place in the cartel period (taking the value of 1 for a cartel and 0 otherwise). Moreover, it includes a categorical variable for the contract type (taking the value of 1 for contracts in road construction and asphalting, 2 for mixed contracts, including road construction and civil engineering and 3 for civil engineering contracts), as well as the anonymized date and year. The first year in our sample begins with a value of 1 and the last year ends with a value of 14. The first anonymized date equals 42, and the last 4,886. To ensure anonymization of the bids, COMCO multiplied them with a factor between 1 and 1.2. The transformation does not affect the calculation of the screens.

Table 2 provides key information on the Swiss data. In order to calculate the predictors of our empirical analysis, we consider tenders with four bids or more. In total, there are 310 tenders with complete cartels with a total value of more than 110 million CHF and 2,031 bids submitted by the cartel participants. Furthermore, there are 287 tenders with incomplete cartels with a total value of more than 114 million CHF. Cartel participants submitted 1,414 bids in these tenders and external firms 650 bids. Finally, we observe 2,398 competitive tenders with a value of roughly 1,700 million CHF and 13,925 submitted bids. In Appendix D, we present additional descriptive statistics of the Swiss data.

3 Detection Methods

This section outlines our novel approach to detect bid rigging. We first describe the concept of a random forest, the machine learning algorithm used for training and testing predictive models for collusion (see Ho 1995; Breiman 2001). Second, we present the screens that enter in the algorithm as potential predictors in detail. Third, we discuss five different predictive models applied to our data that differ in the

[19] See press release: https://www.newsd.admin.ch/newsd/message/attachments/58229.pdf.

[20] See decision *Strassenbau Graubünden*, R 197.

[21] https://www.admin.ch/opc/en/classified-compilation/20022540/index.html.

Table 2 Overview for the Swiss data

Tenders with complete cartels	310
Volume of tenders with complete cartels in million CHF	111.74
Collusive bids in tenders with complete cartels	2,031
Tenders with incomplete cartels	287
Volume of tenders with incomplete cartels in million CHF	114.73
Competitive bids in tenders with incomplete cartels	650
Collusive bids in tenders with incomplete cartels	1,414
Competitive tenders	2,398
Volume of competitive tenders	1,735.91
Competitive bids in competitive tenders	13,925
Tenders for road construction and asphalting	1,389
Tenders for civil engineering	1,286
Tenders for mixed contracts	273

included screens. Finally, we provide descriptive statistics for two important screens in each dataset.

3.1 Random Forest

We use the random forest as a machine learning algorithm for predicting collusive and competitive tenders. In our data, the outcome is given a value of 1 for collusive tenders, including both incomplete and complete bid-rigging cartels, and 0 for competitive tenders. Note that we intentionally do not distinguish between incomplete and complete cartels, as we aim to construct a reliable method for detecting any form of bid rigging. Tenders are therefore either collusive or competitive.

Machine learning requires the data to be randomly split into the so-called training data, used to develop the predictive model, and the test data used to evaluate the model's performance. We randomly split the data such that the training and test data consist of 75 and 25% of the observations, respectively. The random forest is a so-called ensemble method that averages over multiple decision trees to predict the outcome. Tree-based methods split the predictor space (according to the values, which the screens might take) of the training data recursively into a number of non-overlapping regions. Each split aims to maximize the homogeneity of the dependent variable within the newly created regions according to a goodness of fit criterion like the Gini coefficient. The latter measures the average gain in purity (or homogeneity) of outcome values when splitting and is popular for binary variables like our collusion dummy. Splitting is continued until the decision tree reaches a specific stopping rule, e.g. a minimum number of observations in a region or a maximum number of splits. Tree-based predictions of bid rigging (1) or competition (0) are based on whether collusive or competitive tenders dominate in the region that contains the values of the screens for which the outcome is to be predicted.

Importantly, there exists a bias-variance trade-off in out of (training) sample prediction when using such tree-based (and other machine learning) methods when it comes to model generality. More splits reduce the bias and increase the flexibility of the model specification, though at the cost of a greater variance in the unseen data, as the test sample is not used for training due to the regions being smaller. The issue of a too large variance can be mitigated by repeatedly drawing many subsamples from the initial training data and estimating the predictive model, *i.e.* the tree (or splitting) structure, in each of the newly generated samples. For this reason, we apply a random forest algorithm, which predicts the collusion outcome by a majority rule based on the individual trees. This means that the outcome is classified as collusion or competition depending on whether the majority of the trees estimated in the various subsamples predicts collusion or competition, respectively, for particular values of the screens. A further feature of the random forest is that at each splitting step in a specific subsample, only a random subsample of possible predictors (i.e. screens) is considered, reducing the correlation of tree structures across the subsamples and thus further reducing the prediction variance. In our application, we use the randomForest package by Breiman and Cutler (2018) for the statistical software R with growing 1,000 trees to estimate the predictive models in the training data and assess their performance in the test data based on the correct classification rate.

Note that we repeat the random sample splitting into 75% training and 25% test data and assess the predictive performance in the latter 100 times. Our reported correct classification rate corresponds to the average of the correct classification rates across the 100 repetitions. This procedure is likely to entail a smaller variance in estimating the correct classification rate than relying on a single random data split.

3.2 Screens

Screens are statistics that permit data analysis to flag anomalous outcomes indicating potential anticompetitive issues. The literature usually differentiates structural from behavioral screens in cartel detection (see Harrington 2008; OECD 2013; Froeb et al. 2014). Structural screens focus on the factors facilitating the emergence of collusive agreements and help to identify markets in which collusion is more likely. Among these factors, distinctions are made between market structure, demand-related factors, and supply-related factors (OECD 2013). In contrast, behavioral screens empirically measure the behavior of market participants and assess whether the observed behavior significantly departs from competitive behavior to flag it as a potential issue worth scrutinizing further. Following Huber and Imhof (2019) we propose using various descriptive statistics as screens and combining them with machine learning, however, to uncover not only complete but also incomplete bid-rigging cartels.[22] We consider three classes of screens: variance, asymmetry, and uniformity.

[22] In contrast to the context of causal inference, causality goes from the dependent variable (collusion or competition) to the predictors (screens) rather than the other way around. The incidence of collusion as an explanatory variable affects the distribution of bids and thus the screens in causal terms. As in Huber and Imhof (2019) our prediction problem consists of analyzing a reverse causality. By investigating the screens, we infer the existence of their cause: collusion.

As variance screens, we implement the coefficient of variation (CV) and the kurtosis statistic (KURTO), as suggested by Huber and Imhof (2019) and Imhof (2019). In addition, we also implement the spread (SPD) of the distribution of the bids as a screen.

The coefficient of variation is widely discussed in the literature (see Abrantes-Metz et al. 2006; Esposito and Ferrero 2006; Jimenez and Perdiguero 2012; Abrantes-Metz et al. 2012; Imhof 2019) and is defined as the standard deviation divided by the arithmetic mean of all bids submitted in a tender:

$$CV_t = \frac{s_t}{\bar{b}_t}, \tag{1}$$

where s_t is the standard deviation and $\bar{b}_t$ is the mean of the bids in some tender t. The coordination and manipulation of bids by cartel participants might affect the convergence in the distribution of the bids. More precisely, we suspect that bids converge when firms in an auction form a cartel. This is the case because cover bids are somewhat higher than that of the designated winner and concentrate around similar values which are considered to be large enough to make sure that the lowest bid wins. For this reason, the following kurtosis statistic appears appropriate for capturing such convergence effects in cover bids:

$$KURTO_t = \frac{n_t(n_t+1)}{(n_t-1)(n_t-2)(n_t-3)} \sum_{i=1}^{n_t} \left(\frac{b_{it} - \bar{b}_t}{s_t}\right)^4 - \frac{3(n_t-1)^3}{(n_t-2)(n_t-3)}, \tag{2}$$

where b_{it} denotes the bid i in tender t, n_t the number of bids in tender t, s_t the standard deviation of bids, and $\bar{b}_t$ the mean of bids in that tender. To put it bluntly, the smaller the difference between the bids, the higher the kurtosis statistic, and thus, the higher the incidence for a collusive situation. Furthermore, we estimate the spread using the following formula:

$$SPD_t = \frac{b_{max,t} - b_{min,t}}{b_{min,t}}, \tag{3}$$

where $b_{max,t}$ denotes the maximum bid and $b_{min,t}$ the minimum bid in some tender t.

As bid rigging may produce asymmetries in the distribution of bids, we implement the following cover-bidding screens as in Huber and Imhof (2019): the percentage difference (DIFFP), the skewness (SKEW), the relative distance (RD), and the normalized distance (RDNOR). In addition, we add an alternative measure for calculating the relative difference, namely the alternative relative distance (RDALT).

It seems plausible that cartel participants manipulate the difference between the lowest and second lowest bids to secure that the contract is awarded to the cartel's designated winner. To analyze the difference between the two lowest bids, we use the following formula to calculate the percentage difference:

$$DIFFP_t = \frac{b_{2t} - b_{1t}}{b_{1t}}, \tag{4}$$

where b_{1t} is the lowest bid and b_{2t} the second-lowest bid in some tender t. We also

consider the absolute difference between the first and second-lowest bids $D_t = b_{2t} - b_{1t}$ in the empirical analysis.

The manipulation of bids by cartel participants can simultaneously affect both the difference between the first and the second-lowest bid and the differences across the losing bids. Therefore, following Imhof et al. (2018), we calculate a relative distance (relative to a measure of dispersion) in a tender by dividing the difference between the first and the second-lowest bid by the standard deviation of the losing bids:

$$RD_t = \frac{b_{2t} - b_{1t}}{s_{losingbids,t}}, \tag{5}$$

where b_{1t} denotes the lowest bid, b_{2t} the second-lowest bid, and $s_{t,losingbids}$ the standard deviation calculated among the losing bids in some tender t. In terms of its predictive power the RD was outperformed by the difference between the first and the second-lowest bid divided (or normalized) by the average of the differences between all adjacent bids (see Huber and Imhof 2019)). We also consider this normalized distance in our study:

$$RDNOR_t = \frac{b_{2t} - b_{1t}}{\frac{(\sum_{i=1,j=i+1}^{n_t-1} b_{jt} - b_{it})}{n_t - 1}}, \tag{6}$$

where b_{1t} is the lowest bid, b_{2t} the second-lowest bid, n_t is the number of bids and b_{it}, b_{jt} are adjacent bids (in terms of price) in tender t, with bids being arranged in increasing order.

We consider a further alternative measure for the relative distance, initially suggested by Imhof et al. (2018):

$$RDALT_t = \frac{b_{2t} - b_{1t}}{\frac{(\sum_{i=2,j=i+1}^{n_t-1} b_{jt} - b_{it})}{n_t - 2}}, \tag{7}$$

where b_{1t} is the lowest bid, b_{2t} the second-lowest bid, n_t is the number of bids and b_{it}, b_{jt} are adjacent losing bids in a tender t, with bids being arranged in increasing order. In contrast to the normalized distance, the mean of the differences in the denominator is calculated using only losing bids. Furthermore, bid manipulation might affect the symmetry of the distribution of bids. For example, due to a greater difference between the first and the second-lowest bid, we assume that bid-rigging causes a more asymmetric distribution of bids. We, therefore, include the skewness as a screen:

$$SKEW_t = \frac{n_t}{(n_t - 1)(n_t - 2)} \sum_{i=1}^{n_t} \left(\frac{b_{it} - \bar{b}_t}{s_t}\right)^3, \tag{8}$$

where n_t denotes the number of the bids, b_{it} the ith bid, s_t the standard deviation of the bids, and $\bar{b}_t$ the mean of the bids in tender t.

Finally, we verify whether bid rigging (or competition) transforms the distribution of the bids into a less uniform distribution. More concisely, we again suspect that the

higher difference between the first and the second-lowest bid influences the asymmetry, and thus a cartel leads to a less uniform distribution. Therefore, we consider the nonparametric Kolmogorov–Smirnov statistic (KS):

$$D_t^+ = max_i(x_{it} - \frac{i_t}{n_t + 1}), D_t^- = max_i(\frac{i_t}{n_t + 1} - x_{it}), KS_t = max(D_t^+, D_t^-), \quad (9)$$

where n_t is the number of bids in a tender, i_t the rank of a bid and x_{it} the standardized bid for the ith rank in tender t. The standardized bids x_{it} are the bids b_{it} divided by the standard deviation of bids in tender t to facilitate the comparison of tenders with different contract values. We suspect that the KS-statistic should generally differ across cartels and competitive periods.

3.3 Summary Screens

In incomplete cartels, competitive bidders distort the statistical signals produced by bid rigging in the distribution of bids in a tender. We demonstrate this in Fig. 1. Suppose we have four colluding firms colluding. We would assume that the bids converge when firms form a cartel, which could e.g. be detected based on a reduced coefficient of variation, as exemplified in the top-left panel of Fig. 1. However, a competitive bidder might distort the statistical signal produced by the cartel if bidding (significantly) lower or higher than cartel members, as exemplified by bidder 5 in the bottom-left and the top-right panel of Fig. 1. Only if the competitor submits a bid close to the collusive bids, the signal will remain (almost) unaffected, as it is the case in the bottom-right panel. Such a situation can result from a competitor trying to enjoy the umbrella effect and bidding closer to the collusive bids or from pure coincidence.

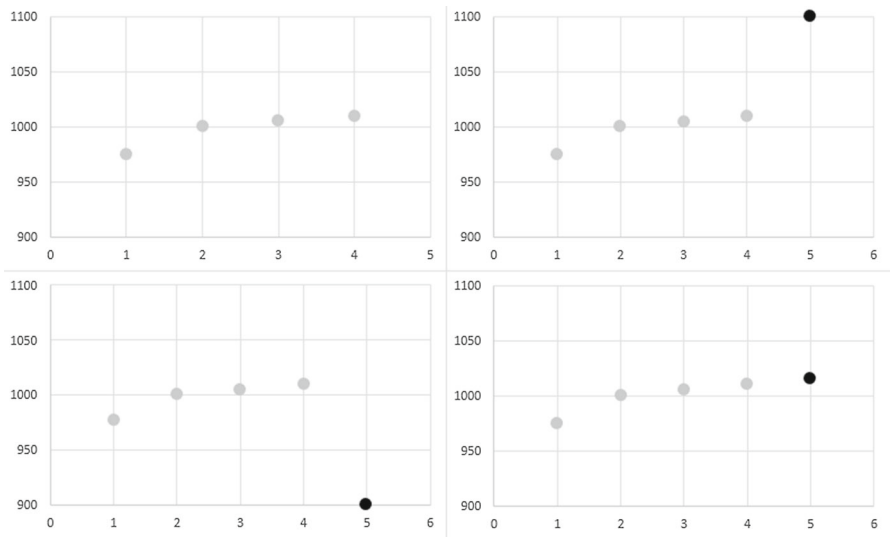

Fig. 1 The potential effect of a competitive bidder

Table 3 Example of possible subgroups for three and four bids in a tender

Bids in a tender	Subgroups formed with three bids	Subgroups formed with four bids
4	4	1
5	10	5
6	20	15
7	35	35
8	56	70
9	84	126
10	120	210

Therefore, the tender-based screens can fail to recognize bid rigging if they are calculated for all bids. We circumvent that distortion by calculating the screens not (only) for all the bids in a tender but all possible subgroups of three and four bids. Table 3 gives the number of possible subgroups of three or four bids, respectively, when the total number of bids in a tender varies between four to ten bids.

For instance, in a tender with a total number of six bids, we calculate the same screen but for 15 different subgroups containing four bids and for 20 different subgroups containing three bids. In each tender, we then include summary statistics for each screen: the mean, the median, the minimum, and the maximum of the respective screen across the various subgroups of three or four bids. We use these summary statistics, so-called 'summary screens', as predictors for flagging collusive and competitive tenders and, therefore, permit comparing tenders with different numbers of bids. We subsequently exemplify the computation of such summary screens by means of the coefficient of variation for subgroups formed on four bids.

The mean of all coefficients of variation calculated for subgroups of four bids in each tender is:

$$MEAN4CV_t = \sum_{s=1}^{N_t} \left(\frac{s_{st}/\bar{b}_{st}}{N_t}\right), \tag{10}$$

where s and t denote the indices for some sub-group s and some tender t respectively, N_t is the number of all the possible subgroups of four bids in tender t and s_{st} and $\bar{b}_t$ are the standard deviation and the mean of the bids respectively. Likewise, the minimum and maximum of the coefficients of variation across the subgroups in a tender correspond respectively to:

$$MIN4CV_t = min_s \frac{s_{st}}{\bar{b}_{st}}, \tag{11}$$

$$MAX4CV_t = max_s \frac{s_{st}}{\bar{b}_{st}}, \tag{12}$$

In order to calculate the median for subgroups of four bids in each tender, define the

coefficient of variation in subgroup s and tender t as $CV_{st} = \frac{s_{st}}{\bar{b}_{st}}$ and order the coefficients in so that

$$CV_{1t} \leq CV_{2t} \leq \ldots \leq CV_{st} \leq \ldots \leq CV_{N_t t}.$$

If the number of subgroups N_t in a tender is uneven, the median of the coefficient of variation in tender t is calculated as follows:

$$MEDIAN4CV_t = CV_{(N_t+1)/2,t}, \tag{13}$$

If the number of subgroups is even, the median corresponds to:

$$MEDIAN4CV_t = \frac{CV_{N_t/2,t} + CV_{N_t/2+1,t}}{2}. \tag{14}$$

We apply these approaches to all the screens discussed above across the different tenders. Note that we do not calculate summary screens for subgroups of two bidders because of the impossibility of calculating some screens as the relative distance (RD), the alternative measure for the relative distance (RDALT), the normalized distance (RDNOR), the kurtosis statistic (KURTO), or the skewness (SKEW). Moreover, cartel participants usually numbered more than two in tenders characterized by incomplete cartels. We also renounce calculating screens for sub-groups of five bidders or more. Using summary screens calculated for subgroups of five bidders only makes sense for tenders with six bids and more. Using tenders with six bids or more would have restricted our sample too much and limited the application of our suggested methods in other cases. Finally, our original application of summary screens does not require the identity of bidders. Instead, we only need the bids in each tender to apply them in many different contexts.

Appendix E presents the descriptive statistics for the samples used in the empirical analyses of the Swiss data.

3.4 Model Specification

In the empirical analyses, we consider five different predictive models that vary in terms of screens considered and a benchmarking method. For the latter, we use the benchmarks suggested by Imhof et al. (2018), developed for and applied to the Swiss construction market.[23] Model 1 only includes screens calculated for all bids in a tender (rather than summary screens). This approach relates to the one discussed by Huber and Imhof (2019). Still, it extends the set of predictors compared to the study by including the relative measure for the alternative distance (RDALT), the spread (SPD), and the Kolmogorov-Smirnov statistic (KS). In total, we use nine predictors and exclude all screens based on the absolute bid value to consider only scale-invariant screens in model 1.

In contrast, model 2 exclusively includes the summary screens, calculated for all possible subgroups of three bids in a tender. In total, we consider the application of

[23] More precisely, tenders with a coefficient of variation (CV) below 6 and a relative distance (RD) above 1 are classified as conspicuous.

32 of these summary screens, using all screens of model 1 but the kurtosis (KURTO), which requires at least four bids. Model 3 uses the summary screens of all screens presented above for all possible subgroups of four bids in a tender, making a total of 36 predictors (now including the kurtosis). Model 4 considers all predictors included in models 1, 2, and 3, resulting in 77 screens in total, mixing the summary screens with the tender-based screens. Finally, model 5 also includes three screens based on absolute bid values (and thus not scale-invariant) and the number of bids in a tender (NBRBIDS), producing 81 predictors in total. The motivation for including the number of bids is that it might be easier to settle an agreement with fewer rather than more bidders. Moreover, we can account for behavioral responses of bidders to fiercer competition due to an increased number of bidders (see, e.g., Vickrey 1961). The three value-based screens are the mean bid in a tender included as a proxy for the contract value (MEANBIDS), the standard deviation of the bids in a tender (STDBIDS), and the absolute difference between the first and the second-lowest bid (D).

4 Flagging Incomplete Bid-Rigging Cartels

4.1 Application

We apply our detection method to data drawn from the cases of See-Gaster and Strassenbau Graubünden, characterized by well-organized bid-rigging cartels, which, however, faced competitive outsiders from time to time. In these real cases, competitive and collusive bidders were aware of their reciprocal existence. Evidence from COMCO's investigations has pointed out that cartel participants adopted a more competitive behavior in the presence of competitive bidders by deciding not to collude in some tenders. An agreement's poor chance of success due to several (potential) competitive bidders motivated such decisions to bid independently for some contracts. However, in other tenders, the cartel faced only one competitive bidder and tried to include her or him in the agreement. Moreover, competitive bidders aware of the existence of the cartel might have tried—if not enrolled in the agreement—to benefit from the umbrella effect of a cartel by bidding higher than they would have in a fully competitive situation. As a consequence of the umbrella effect, bids of the competitive bidder fall nearer to bids of collusive bidders, such as competitive bids distort less the statistical pattern produced by bid rigging (as illustrated in the bottom-right panel of Fig. 1).

We construct different samples of collusive tenders. Sample 1 includes all tenders with incomplete bid-rigging cartels and at least two cartel participants. As stated in Table 4, the average percentage of cartel participants in sample 1 amounts to 71%. Sample 2 includes tenders with incomplete bid-rigging cartels formed with at least three cartel participants. Since sample 2 excludes tenders with only two cartel participants, the average rate of cartel participants of 75% in sample 2 is superior to that in sample 1. The logic is the same for samples 3 to 5. Consequently, sample 5 has the highest average percentage of cartel participants and contains fewer

Table 4 Correct classification rate in the Swiss data

Sample	Cart.F	Perc.Cart.F	Tenders	Bench.	M1	M2	M3	M4	M5
1	> 1		All	0.524	0.612	0.637	0.642	0.645	0.673
			Comp.	0.901	0.612	0.607	0.612	0.62	0.646
			Coll.	0.147	0.615	0.671	0.677	0.677	0.706
2	> 2	75%	All	0.525	0.648	0.665	0.675	0.678	0.708
			Comp.	0.901	0.652	0.643	0.645	0.65	0.683
			Coll.	0.148	0.647	0.691	0.71	0.709	0.737
3	> 3	79%	All	0.511	0.706	0.722	0.748	0.745	0.759
			Comp.	0.901	0.705	0.688	0.705	0.707	0.719
			Coll.	0.121	0.708	0.758	0.792	0.784	0.800
4	> 4	83%	All	0.506	0.743	0.770	0.8	0.798	0.814
			Comp.	0.901	0.755	0.751	0.764	0.771	0.783
			Coll.	0.111	0.735	0.791	0.835	0.826	0.846
5	> 5	88%	All	0.494	0.766	0.805	0.813	0.818	0.841
			Comp.	0.901	0.771	0.769	0.786	0.788	0.813
			Coll.	0.087	0.763	0.844	0.842	0.849	0.871
Compl. Cartel	All	100%	All	0.617	0.826	0.813	0.82	0.823	0.833
			Comp.	0.900	0.83	0.818	0.819	0.827	0.833
			Coll.	0.334	0.823	0.808	0.823	0.82	0.834

Note: "Sample", "Cartel.F", "Per.Cart.F", "Tenders", "Bench.", "M1", "M2", "M3", "M4" and "M5" denote the sample, the number of cartel firms in the collusive tenders, the percentage of cartel firms in the collusive tenders, the type of tenders, the results produced by the screening methods of the benchmarking approach, model 1, model 2, model 3, model 4 and model 5 respectively. For the type of tenders, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively

competitive bidders but at least one per tender. In addition, we construct a sample including all tenders with complete cartels.

First, we investigate the performance of predictive models starting with complete cartels. As shown in Table 4, the correct classification rates do not differ notably across machine learning-based models 1 to 5, the range being 81.3% to 83.3%. However, for the benchmarking method, the correct classification rate of 61.7% is clearly below that of models 1 to 5. In addition, it differs strongly between competitive and collusive tenders, amounting to only 33.4% in the latter case. Possible explanations for this poor performance are the reliance on only two screens, which are not necessarily the optimal predictors, and the use of benchmark values for these two screens drawn from two previous investigations, which are not necessarily optimal in the dataset being considered. In contrast, machine learning approaches use a more extensive set of screens and weight their importance in a data-driven way.

However, if we adjust the benchmarks of our benchmarking approach applied, we can achieve better prediction rates for complete cartels. In Appendix B, we depict a decision tree on Fig. 2 corresponding to the minimal cross-validation error. Our pruned tree, using as predictors only the relative distance (RD) and the coefficient of variation (CV) as in Imhof et al. (2018), shows a correct classification rate of 81.6% for complete cartels. This discrepancy illustrates the fundamental difference between a benchmark method and machine learning: benchmarks are exogenous, whereas machine learning outperforms benchmarks since it chooses the best predictors in each case. While a benchmark can still be adapted to different cases, machine learning algorithms are far more precise. Nonetheless, a benchmark method requires less information to be implemented and therefore remains a simple (first) step in flagging cartels.

Considering models 1 to 5, the correct classification rates vary between 61.2% and 84.1%, depending on the sample and the model. When the proportion of competitive bidders increases, the correct predictions generally decrease, as depicted in Table 4. This result suggests that cartel participants anticipated competitive bids and decided not to collude in some peculiar tenders, for example, in the case of See-Gaster. The models with summary screens calculated for subgroups outperform model 1. Among them, models 3 and 4 slightly outperform model 2, indicating that in our case, summary screens calculated for subgroups of four bids exhibit a higher predictive power than those calculated for subgroups of three bids. The fact that we have four cartel participants per tender in most cases likely explains this result. In contrast, summary screens calculated for subgroups of three bids may work better if we mainly observe three cartel participants per tender.

Model 5, the only one including the number of bidders or the contract value as predictors, outperforms the other models and has correct classification rates of 5 to 10 percentage points higher than model 1. The advantage of models 3 or 4 over model 1 varies from 3 to 5.7 percentage points. This points to a decrease in the error rate by roughly more than 20% in some cases, even in the presence of potentially strategic interactions, i.e., outsiders aware of the existence of bid-rigging cartels and trying to benefit from the umbrella effect (Bos and Harrington 2010). Therefore, competition agencies should consider summary screens for subgroups to detect both complete and incomplete bid-rigging cartels.

Moreover, note that the benchmarking method poorly performs when flagging incomplete bid-rigging cartels and does no better than tossing a coin. Specifically for truly collusive tenders, the correct classification rates vary only between 8.7% and 14.7%.

When looking at the variable importance as reported in Table 5, we find for all models and samples that the Kolmogorov-Smirnov statistic (KS) is an important predictor. In many cases, it is among the three most important variables. This suggests that even if collusive and competitive tenders generally do not follow a uniform distribution, collusive bids are usually far less uniform than competitive bids. Therefore, the Kolmogorov-Smirnov statistic for deviations from the uniform distribution tends to exhibit notably higher values in rigged tenders than in competitive tenders.

The random forest generally picks up a balanced set of screens for the variance and asymmetry along with the Kolmogorov-Smirnov statistic for model 1 in all samples. Specifically for the sample with complete cartels, we observe for models 2 to 5 that the random forest selects screens for the variance, mainly the coefficient of variation (CV) and the spread (SPD), along with the Kolmogorov-Smirnov statistic (KS). Screens for asymmetry in the distribution of bids remain unselected for models 2 and 5 when the cartel is complete. However, when cartels are incomplete, the random forest selects for models 2 to 5 screens for asymmetry in the distribution of bids, mostly the skewness (SKEW), the relative distance (RD), the percentage difference (DIFFP), and the alternative distance (RDALT). Even though the results suggest that the screens for asymmetry are less important than the screens for variance and the Kolmogorov-Smirnov statistic (KS).

For all samples with incomplete cartels, the minima and maxima of the summary screens are the most important predictors, while the mean and median are most important in complete cartels. The results suggest that a few competitive bids sufficiently disturb the statistical pattern produced by bid rigging that it becomes difficult to detect collusion using tender-based screens. In contrast, the use of the minimum or maximum of summary screens mitigates the distortion of competitive bids in the statistical patterns produced by bid rigging and it allows us to detect with a high probability both incomplete and complete bid-rigging cartels in the Swiss data.

4.2 Robustness Analysis

We investigate the robustness of our results by discarding the most important predictors and applying the random forest to the remaining predictors. Since model 1 uses fewer predictors than the other models, we leave out the three most important variables, while for models 2 to 5, we drop the five best predictors. Table 6 reports the difference in percentage points in the correct classification rates when keeping vs. dropping the respective predictors.

The overall correct classification rate of model 1 in samples 1, 3, and 4, keeping all variables, predominates when dropping the three best predictors by 3.4 to 4.8 percentage points. Considering the other models and samples, we observe more or less the same predictive power when discarding the most important variables. Therefore, the remaining predictors seem to be suitable substitutes for the discarded ones. Other variables become more important when the most important predictors are omitted, and the correct classification rate is hardly affected.

Furthermore, we investigate the robustness for the type of contract. We subsequently only consider contracts for road construction and asphalting for both the cartel and post-cartel periods. We exclude contracts for civil engineering and mixed contracts combining civil engineering with road construction or asphalting. The reason for this is that certain specific characteristics of contracts in civil engineering might affect the screens and, therefore, the correct classification rate. Dropping mixed contracts and contracts for civil engineering permits us to verify whether this importantly affects the correct classification rate among the remaining contracts for road construction and asphalting. Table 7 reports the difference in

Table 5 Important predictors for the Swiss data

	Model 1		Model 2		Model 3		Model 4		Model 5	
	IV	MDG	IV	MDG	IV	MDG	IV	MDG	IV	MDG
1	SKEW	22.48	MIN3CV	9.98	MIN4CV	9.87	MIN4CV	5.8	STDBIDS	5.94
	RDNOR	21.73	MAX3KS	9.93	MAX4KS	9.55	MAX4KS	5.56	MEANBIDS	5.67
	SPD	21.67	MIN3SPD	9.66	MIN4SPD	9.34	MIN4SPD	5.39	MIN4CV	5.27
	KS	21.31	MAX3DIFFP	6.56	MIN4SKEW	8.14	MAX4RD	4.66	MAX4KS	5.11
	DIFFP	20.75	MIN3DIFFP	6.47	MAX4RD	8.08	MIN4SKEW	4.61	MIN4SPD	4.92
2	KS	19.64	MIN3CV	9.27	MIN4CV	9.26	MIN4CV	5.61	MIN4CV	5.23
	RDNOR	19.56	MAX3KS	9.23	MAX4KS	9.12	MAX4KS	5.54	STDBIDS	5.03
	SPD	19.51	MIN3SPD	8.98	MIN4SPD	8.87	MIN4SPD	5.32	MAX4KS	5.02
	SKEW	18.66	MIN3DIFFP	5.62	MAX4RD	7.71	MAX4RD	4.63	MEANBIDS	4.97
	CV	18.38	MAX3DIFFP	5.51	MIN4SKEW	7.67	MIN4SKEW	4.56	MIN4SPD	4.78
3	RDNOR	16.31	MAX3KS	8.42	MIN4CV	8.32	MIN4CV	5.46	NBRBIDS	5.88
	SPD	15.04	MIN3CV	8.32	MAX4KS	8.08	MAX4KS	5.43	MIN4CV	4.9
	KURTO	14.62	MIN3SPD	8.2	MIN4SPD	7.52	MAX4RD	5.01	MAX4KS	4.87
	KS	14.35	MAX3RD	5.09	MAX4RD	7.5	MIN4SPD	4.96	MAX4RD	4.53
	RDALT	14.02	MAX3RDALT	5.09	MIN4SKEW	7.4	MIN4SKEW	4.94	MIN4SKEW	4.49
4	RDNOR	13.28	MIN3SPD	8.15	MIN4CV	7.6	MIN4CV	5.15	NBRBIDS	7.49
	SPD	12.12	MAX3KS	8.13	MAX4KS	7.49	MAX4KS	5.09	MIN4CV	4.52
	KS	11.52	MIN3CV	8.09	MIN4SPD	7.46	MIN4SPD	4.99	MAX4KS	4.45
	RDALT	11.42	MAX3RD	4.51	MIN4SKEW	7.39	MIN4SKEW	4.96	MIN4SKEW	4.38
	CV	11.06	MIN3SKEW	4.49	MAX4RD	6.92	MAX4RD	4.66	MIN4SPD	4.33

Table 5 continued

	Model 1		Model 2		Model 3		Model 4		Model 5	
	IV	MDG	IV	MDG	IV	MDG	IV	MDG	IV	MDG
5	KS	10.35	MIN3SPD	6.43	MAX4KS	6.05	MIN4CV	4.08	NBRBIDS	7.27
	SPD	10.14	MIN3CV	5.98	MIN4CV	6.05	MAX4KS	3.99	MIN4CV	3.51
	CV	9.83	MAX3KS	5.94	MIN4SPD	6.02	MIN4SPD	3.96	MIN4SPD	3.4
	RDNOR	9.53	MAX3RD	3.75	MIN4SKEW	5.09	MAX4RDALT	3.25	MAX4KS	3.39
	RDALT	8.03	MAX3RDALT	3.74	MAX4RDALT	4.96	MAX4RDNOR	3.25	MAX4RDALT	2.86
Compl. Cartel	KS	54.65	MEDIAN3CV	20.04	MEDIAN4KS	20.23	MEDIAN4KS	11.4	MEDIAN4KS	10.86
	CV	50.9	MEDIAN3SPD	18.68	MEDIAN4SPD	20.04	MEDIAN4SPD	11.27	MEDIAN4CV	10.82
	SPD	32.96	MEAN3CV	18.49	MEDIAN4CV	19.56	MEDIAN4CV	11.04	MEDIAN4SPD	10.75
	DIFFP	18.37	MEDIAN3KS	18.46	MEAN4KS	16.91	MEDIAN3CV	10.29	MEDIAN3CV	9.96
	RDNOR	17.1	MEAN3SPD	17.64	MEAN4CV	16.67	MEDIAN3KS	9.37	MEDIAN3SPD	9.22

Note: “IV” and “MDG” denote the important variables selected by the random forest and the mean decrease in Gini index. ”KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, “DIFFP” and ”KURTO” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference and the kurtosis statistic respectively. For the abbreviations, see also Table 9 in Appendix A.

Table 6 Differences between original random forest and random forest with discarded variables

Sample	Cart. F	Perc. Cart.F	Tenders	M1	M2	M3	M4	M5
1	> 1	71%	All	3.4	1.1	-0.2	0	1.8
			Comp.	3.6	-0.9	-1.2	-0.3	1.8
			Coll.	3.2	3.1	0.9	0.3	1.8
2	> 2	75%	All	1.8	1.3	-0.4	0	1.4
			Comp.	2.9	-0.4	-1.1	-0.3	1.4
			Coll.	0.6	3	0.3	0.3	1.3
3	> 3	79%	All	4.8	0.1	0.1	0.1	0.5
			Comp.	6.3	-0.4	-0.2	-0.2	0.1
			Coll.	3.2	0.6	0.6	0.5	0.9
4	> 4	83%	All	3.4	1.2	0.2	0.9	1.8
			Comp.	4.1	0.9	-0.2	0.5	0.9
			Coll.	2.6	1.6	0.6	1.3	2.7
5	> 5	88%	All	0.6	2	0.2	0.2	1.7
			Comp.	1.3	0.9	0.5	0.2	1.5
			Coll.	-0.2	3.1	-0.4	0	1.7
Compl. Cartel	All	100%	All	1	-0.2	0.4	0	0
			Comp.	0.7	0.5	0.6	0.2	-0.2
			Coll.	1.3	-0.8	0.3	-0.1	0.2

Note: "Sample", "Cartel.F", "Per.Cart.F", "Tenders", "M1", "M2", "M3", "M4" and "M5" denote the sample, the number of cartel firms in the collusive tenders, the percentage of cartel firms in the collusive tenders, the type of tenders, model 1, model 2, model 3, model 4 and model 5 respectively. For the type of tenders, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively

percentage points in the correct classification rates when using all contracts vs. using contracts for road construction and asphalting only.

In samples 1 and 2, we find the correct classification rates of the random forest for road construction and asphalting contracts to be superior to the classification rate of the random forest with all types of contracts. For example, the difference in the (overall) classification rate of model 1 in samples 1 and 2 accounts for 6.2 and 2.8 percentage points, respectively. A possible explanation could be that we implicitly suppress some competitors when we keep only the road construction and asphalting contracts. For example, in sample 1, the average percentage of collusive bidders is 81%, which is considerably higher, as is the situation with all types of contracts (71%, see Table 4). Therefore, the cartel percentage is higher for this restricted sample of road construction and asphalting contracts alone and explains the higher performance in samples 1 and 2. In sample 3, the situation begins to change for both

Table 7 Differences between original random forest and random forest using only contracts for road construction and asphalting

Sample	Cart. F	Perc. Cart.F	Tenders	M1	M2	M3	M4	M5
1	> 1	81%	All	-6.2	-5.2	-6.7	-5.5	-3.6
			Comp.	-7.2	-4.6	-6.9	-5.7	-3.8
			Coll.	-5.2	-5.8	-6.2	-5.2	-3.3
2	> 2	82%	All	-2.8	-3	-4.3	-3.6	-1.2
			Comp.	-3.6	-3.1	-4.7	-3.9	-1.4
			Coll.	-1.9	-2.9	-3.8	-3.2	-1
3	> 3	84%	All	0.1	-0.6	-0.9	-0.6	0.2
			Comp.	-1.9	-1.5	-2.7	-2.1	-1.9
			Coll.	2	0	0.5	0.5	1.9
4	> 4	86%	All	1.5	2.2	2	2.4	3.4
			Comp.	2.2	2.7	1.9	2.5	3.1
			Coll.	1	1.9	1.9	2.2	3.6
5	> 5	88%	All	2.6	3.3	2	2.1	2.7
			Comp.	3.1	3.3	2.2	1.8	3
			Coll.	1.9	3.2	1.9	2.4	2.3
Compl. Cartel	All	100%	All	0.7	0.3	0	0.3	0.3
			Comp.	0.9	-0.3	-0.4	0	-0.2
			Coll.	0.4	0.9	0.4	0.5	0.8

Note: "Sample", "Cartel.F", "Per.Cart.F", "Tenders", "M1", "M2", "M3", "M4" and "M5" denote the sample, the number of cartel firms in the collusive tenders, the percentage of cartel firms in the collusive tenders, the type of tenders, model 1, model 2, model 3, model 4 and model 5 respectively. For the outcome classification, "All", "Comp." and "Coll." denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively

types, the correct classification rates being almost identical. Noticeably, the differences increase again for all models in samples 3 and 4. However, not as strong as before and in the opposite direction. Therefore, for an almost identical average percentage of cartel participants, the correct classification rates of the random forest for all types of contracts are slightly superior to those for road construction and asphalting.

To investigate the robustness of the correct classification rate across different machine learning algorithms, we also assess the performance of lasso regression and an ensemble method (including bagged trees, random forest, and neural networks) for all models and samples. We explain these algorithms, also outlined by Huber and Imhof (2019), in more detail in Appendix C. Table 8 reports the difference in

percentage points in the correct classification rates of the random forest minus the correct classification rates of the lasso and ensemble method.

Considering samples 1 and 2 in Table 8, we find that the lasso and ensemble method slightly outperform the random forest. The maximum difference in (overall) correct classification rates across models and samples is 2.9 percentage points. While the somewhat lower rates speak against the random forest, performance is more uniform. Therefore, there is less divergence across both the competitive and collusive periods, which may be important to practitioners. For samples 3, 4 and 5, in general, the lasso and ensemble method slightly outperform the random forest, in two cases even more profoundly, with higher correct classification rates of 4.3 to 6.7 percentage points for model 1 in samples 4 and 5. This implies that in samples 4 and 5 (with a high amount of collusive bidders), considering summary screens does not significantly improve the predictive power of the lasso and ensemble method, in contrast to the random forest. On the other hand, and as for samples 1 and 2, the random forest shows a more uniform performance (e.g. correct classification rates are not too different for competitive and collusive tenders). We find a similar performance with regard to the (overall) correct classification rates between the random forest and the ensemble method for complete cartels. However, the random forest slightly dominates the lasso regression. Considering imbalances in the predictive performance across competitive and collusive periods, the random forest and the ensemble method have a more homogeneous performance than the lasso regression.

To conclude, in Table 8, the random forest shows a somewhat lower correct classification rate than the lasso and the ensemble method. Still, it exhibits a more homogeneous correct classification rate across both the competitive and collusive tenders. All in all, this robustness check shows the stability of our results.

5 Conclusion

In this paper, we have suggested a robust method for flagging bid rigging in tenders that is likely to be more powerful for detecting incomplete cartels than previously suggested methods. Our approach combined screens, i.e, statistics derived from the distribution of bids in a tender, with machine learning to predict the probability of collusion. As a methodological innovation, we calculated the screens for all possible subgroups of three or four bids within a tender and considered summary statistics as the mean, median, maximum, and minimum for each screen as predictors in the machine learning algorithm. By doing so, we improved on the issue that competitive bids may distort the statistical signals produced by bid rigging.

We applied our method to data from the investigations involving incomplete cartels in the regions See-Gaster and Graubünden in Switzerland. The out-of-sample performance of machine learning using summary screens (calculated for all possible subgroups of three and four bids) as predictors outperformed other screening methods. However, the performance of all machine learning-based methods in all models still decreased concerning the relative number of competitive bids in the data

Table 8 Differences between original random and the lasso and the ensemble method

		Sample 1		Sample 2		Sample 3		Sample 4		Sample 5		Compl. Cart.	
		lasso	ens.	lasso	ens.	lasso	ens.	lasso	ens.	lasso	ens.	lasso	ens.
M1	All	-1.5	-2	-1.6	-2.3	-3.7	-2.7	-5.6	-6.7	-6	-4.3	0.9	0.4
	Comp.	-3.5	-1.4	-3.1	-0.7	-4.3	0.4	-5.9	0.9	-4	2.2	4.7	0.6
	Coll.	0.1	-2.9	-0.4	-3.9	-3.1	-5.9	-5.4	-14.1	-8.3	-11	-3	0.1
M2	All	-1.3	-1.1	-0.8	-0.7	-1.8	-1.3	-1.6	-1.8	-0.8	0.4	0.2	-0.8
	Comp.	-13	-9.9	4.8	1.5	2.2	1.4	2.7	1.3	2	2.4	4	-0.8
	Coll.	10.3	7.5	-6.5	-2.9	-5.9	-4	-5.7	-4.7	-3.7	-2	-3.6	-0.9
M3	All	-1.4	-0.9	-1.5	-1.2	-1.5	-1	-0.9	0.2	-0.3	0.6	0.1	0.1
	Comp.	0.9	-0.8	0.5	-0.6	1.9	0.7	0.3	1.9	2.6	-6.1	4.6	1.2
	Coll.	-3.8	-1.1	-3.5	-1.8	-4.9	-2.6	-2.2	-1.7	-3.4	7.5	-4.4	-1`
M4	All	-0.7	-0.7	-0.7	-0.6	-1.4	-0.8	-0.9	-0.5	-0.3	1	0.2	-0.1
	Comp.	2.7	0.3	2.9	0.2	2.2	0.9	1.5	1.2	2.6	2.8	4.7	0.9
	Coll.	-4.3	-1.8	-4.4	-1.5	-5	-2.4	-3.2	-2	-3.4	-0.9	-4.3	-1.1
M5	All	-1.5	-2.9	-1.3	-1.4	-2.5	-1.1	-1.6	-1.8	1	-2.4	0.6	-0.1
	Comp.	-3.4	-3.4	-2.4	-2.2	-2.2	-1.4	-3	-0.1	2.4	5.9	4.4	0.6
	Coll.	0.3	-2.2	-0.1	-0.6	-2.7	-0.8	-0.3	-3.3	-3.5	-13.8	-3.2	-0.9

Note: “lasso”, “ens.”, “M1”, “M2”, “M3”, “M4” and ”M5” denote the lasso, the ensemble of method, model 1, model 2, model 3, model 4 and model 5 respectively. For the type of tenders, ”All”, “Comp.” and “Coll.” denote the prediction for all types of tenders, the prediction for the competitive tenders and the prediction for the collusive tenders respectively

of the investigations involving incomplete cartels. This decrease indicates that cartel participants anticipated competition from non-cartel bidders.

Compared to tender-based screens, summary screens increased the correct classification rate by 3 to 5.7 percentage points for incomplete cartels. This implies a substantial decrease in the error rate (one minus the correct classification rate) of 22.2%, despite the threat that predictive performance might be partially compromised by competitive bidders trying to benefit from the umbrella effect, i.e. bidding closer to collusive bids (Bos and Harrington 2010). As screening by competition agencies can trigger investigations with legal consequences for potential cartel members, such a decrease in the error rate of 22.2% appears highly desirable. Thus, our results demonstrate the usefulness of combining machine learning with an improved set of statistical screens to reduce distortions of competitive bids in incomplete cartels. Moreover, the method appears promising for detecting collusion in other industries or countries.

A limitation of our study is that we restricted ourselves to summary screens calculated per (within a) tender. On the one hand, this makes the method simple to implement on a large scale, i.e., for many tenders, in order to flag those ones appearing suspicious. On the other hand, competition authorities are in a second step required to identify those bidders worth investigating further, e.g., by verifying which bidders participated in multiple tenders among those flagged suspicious. The burden of taking this second step might be overcome by screening methods capable of directly flagging suspicious bidders (rather than tenders) (see, e.g., Imhof and Wallimann 2021). Therefore, combining our proposed summary screens with firm-specific predictors of collusion appears to be a promising agenda for future research.

Appendix

A Nomenclature for Abbreviations

Table 9 Nomenclature for abbreviation

COMCO	Swiss Competition Commission
Swiss data	The data from investigations of See-Gaster and Graubünden
CV	The coefficient of variation
KURTO	The kurtosis statistic
SPD	The spread
DIFFP	The percentage difference
RD	The relative distance
RDNOR	The normalized distance
RDALT	The alternative measure for the relative distance
SKEW	The skewness
KS	The Kolmogorov-Smirnov statistic
MEAN4CV	The mean of all CVs for subgroups of 4 bids
MIN4CV	The minimum of all CVs for subgroups of 4 bids
MAX4CV	The maximum of all CVs for subgroups of 4 bids
MEDIAN4CV	The median of all CVs for subgroups of 4 bids
MEAN3CV	The mean of all CVs for subgroups of 3 bids
MIN3CV	The minimum of all CVs for subgroups of 3 bids
MAX3CV	The maximum of all CVs for subgroups of 3 bids
MEDIAN3CV	The median of all CVs for subgroups of 3 bids

Note: This table exemplifies the summary screens by means of the coefficient of variation. The summary screen "MEAN4CV" indicates that it is the mean (MEAN) of all the coefficients of variation (CV) calculated for subgroups of four bids (4). The same can be interpreted for any screen. For example, the "MAX3KS" in Table 5 for sample 1 and model 2 indicates the maximum (MAX) of the Kolmogorov-Smirnov statistic (KS) for subgroups of three bids (3)

B Classification Tree Adjusting the Benchmarking Rule of Imhof et al. (2018) in the Swiss Data in Sample Only with Complete Cartels

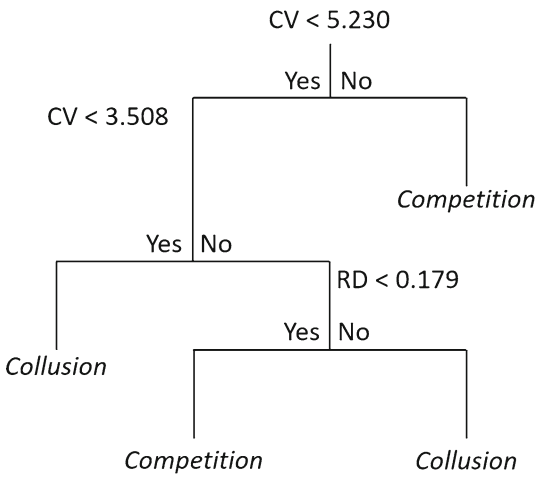

Fig. 2 Adjusted Classification Tree

C Details About Lasso Regression and the Ensemble Method

Here, we discuss the machine learning approaches of the lasso regression and ensemble method in more detail. Similar to the random forest, the lasso regression and ensemble method require randomly splitting the data into training (used for estimating the model parameters) and test data (used for out-of-sample prediction and performance evaluation). Again, our training and test samples contain 75% and 25% of the observations, respectively. Lasso regression corresponds to a penalized logit regression, where the penalty term restricts the sum of the estimated coefficients. The coefficients of less predictive variables thus shrink towards or even exactly to zero depending on the penalty term. Therefore, the lasso regression may perform predictor selection.

The estimation of lasso logit coefficients is based on the following optimization problem:

$$\max_{\beta_0,\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left[y_i\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}\right)-log\left(1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}}\right)\right]-\lambda\sum_{j=1}^{p}|\beta_j|\right\}. \quad (15)$$

where β_0 denotes the intercept, β the estimated coefficients of each predictor, x the vector of predictors, i indexes an observation in our data set (with n being the number of observations), j indexes a predictor (with p being the number of predictors), and λ a penalty term larger than zero. We use the same predictors as described in the main text for the different models. In our application, we use the hdm package by Chernozhukov et al. (2016) for the statistical software R. We apply 15-fold cross-validation to select the penalty term λ based on minimizing the mean squared error of prediction.

For the ensemble method, we apply the "SuperLearner" package for "R" by van der Laan et al. (2008) with default values for bagged regression tree, random forest and neural network algorithms in the "ipredbagg", "cforest" and "nnet" packages respectively. The ensemble method also relies on training data to estimate the model parameters and test data for prediction and performance evaluation. However, any estimation step now consists of a weighted average of bagged classification trees, random forest and neural networks. Bagged trees involve estimating single trees (rather than random forest) repeatedly using the outcome residuals of the respective previous tree as the outcome. Rather than splitting the predictor space, neural networks aim at fitting a system of non-linear functions that models the influence of the predictors of collusion in a flexible way. To do so, we model the association between the predictors and the outcomes using a network of non-linear intermediate functions, so-called hidden notes. Several layers of hidden nodes allow modelling associations and interactions between the predictors in a flexible way. Including more nodes and layers in the network will surely reduce the bias but will also increase the varaince.

D Descriptive Statistics for the Swiss Data

Figure 3 depicts the distribution of the number of bids per tender for complete cartels, incomplete cartels and competitive tenders, respectively. While tenders with four to seven bids dominate, there is also a sufficient number of tenders with eight or more bids (see Table 10).

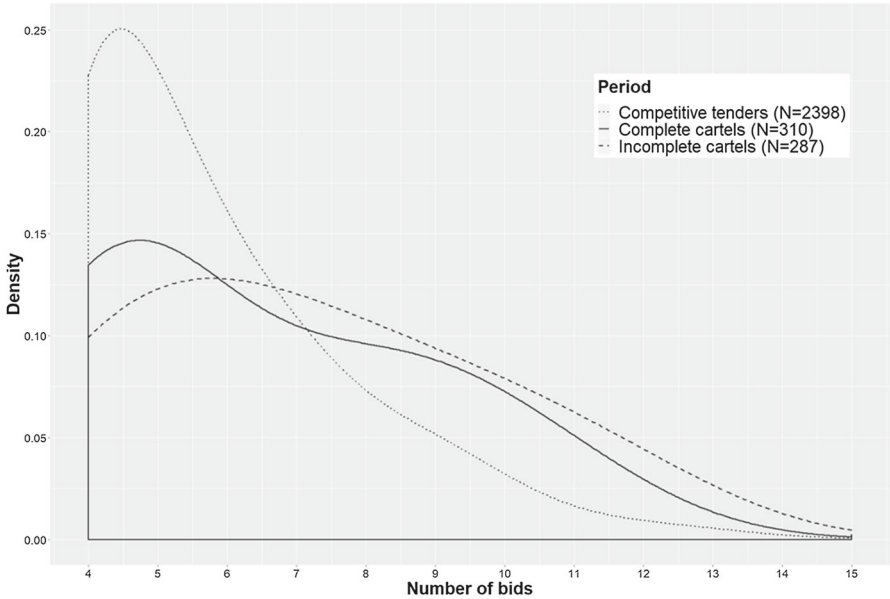

Fig. 3 Distribution of tenders for a predetermined number of bids in the Swiss data

Table 11 depicts the empirical distribution of the lowest bids in a tender with complete and incomplete cartels as well as for tenders with only competitive bids (tenders in the post-cartel period). The empirical distributions of the lowest bids for tenders with complete cartels and with incomplete cartels are similar. However, this is not the case for competitive tenders, which have many more contracts, varying in value from one thousand CHF to 148 million CHF. All the empirical distributions of the lowes bids in a tender are right-skewed, such that the mean is higher than the median. The phenomenon appears to be more noticeable for competitive tenders than for tenders with complete or incomplete cartels.

E Descriptive Statistics for Predictors

In the following tables, we present tables of descriptive statistics for all the different samples used in the empirical analyses. Here, we review the key information drawn from the descriptive statistics for the coefficient of variation (CV) and the normalized distance (RDNOR). Similar interpretations can be made for other screens.

For the Swiss data, we find that the mean of the coefficient of variation during the cartel period amounts to 3.66, the median to 3.29 and the standard deviation to 2.09 (see Table 12). In contrast, the post-cartel period (competitive tenders) in the Swiss data exhibits a mean of 10.12, a median of 8.45 and a standard deviation of 7.89 (see Table 13). Note that in the empirical analyses, we select only competitive tenders with an anonymized year superior or equal to 8. Since collusive tenders superior or equal to 8 are absent in the anonymized years, we conclude that both bid-rigging

Table 10 Numbers of bids in a tender in the Swiss data

Number of bids in a tender	4	5	6	7	8	9	10	10+
Tenders with complete cartels	94	50	29	24	33	33	23	24
Tenders with incomplete cartels	56	36	38	40	27	28	24	38
Competitive tenders	786	559	365	257	158	129	74	70

Table 11 Empirical distributions of bids in the Swiss data (in million CHF)

	Complete cartels	Incomplete cartels	Competitive tenders
Mean	0.36	0.4	0.72
Std	0.36	0.47	3.81
Min	0.03	0.02	0.001
Lower Q.	0.16	0.12	0.13
Median	0.29	0.25	0.31
Upper Q.	0.44	0.50	0.66
Max	3.45	3.46	147.73
N	310	287	2,398

Note: "Mean", "Std", "Min", "Lower Q.", "Median", "Upper Q.", "Max", and "N" denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively

cartels collapsed in this post-cartel period. This ensures that a competitive tender in the post-cartel period is really a "competitive" one.

If we look at the coefficient of variation for the incomplete bid-rigging cartel in sample 1 of the Swiss data (collusive tenders characterized by incomplete cartels with at least two colluding firms), the CV is affected by the presence of competitive bids with a mean of 7.79, a median of 6.79 and a standard deviation of 3.89 (see Table 14). Looking more precisely at the minimum of all coefficients of variation calculated for subgroups of four bids in a tender (MIN4CV), we find a mean of 3.16, a median of 2.26 and a standard deviation of 2.97 for the incomplete bid-rigging cartels in sample 1 (see Table 14). However, the MIN4CV for the competitive tenders exhibits higher values with a mean of 6.24, a median of 4.49 and a standard deviation of 6.77 (see Table 13). Noteworthy, the differences are weaker for the maxima of all coefficients of variation calculated for subgroups of four bids (MAX4CV), between incomplete cartels in sample 1 (mean of 10.63, median of 9.43 and a standard deviation of 5.46 in Table 14) and competitive tenders (mean of 12.14, median of 10.13 and a standard deviation of 9.73 in Table 13). This example is crucial to understand the benefit delivered by summary statistics of the screens. Even if the maxima of the coefficient of variation is high in both cases of incomplete bid-rigging cartels and competition, the minima diverge notably and could be used to differentiate between competition and collusion.

Table 12 Descriptive statistics for the collusive tenders including only cartel participants in the Swiss data

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	6.57	2.44	4	4	6	9	13	308
MEANBIDS	379.88	376.09	34.42	172.13	305.81	460.98	3509.71	308
STDBIDS	13.15	13.67	0.49	5.59	9.75	15.88	109.94	308
CV	3.66	2.09	0.6	2.22	3.29	4.51	15.73	308
KURTO	0.16	1.65	-5.4	-0.99	0.21	1.33	4.37	308
SKEW	0.08	0.81	-1.94	-0.42	0.07	0.7	1.78	308
SPD	0.11	0.07	0.01	0.06	0.09	0.13	0.5	308
D	9.20	13.32	0.14	2.93	6.02	10.11	121.30	308
RD	1.16	1.36	0.01	0.46	0.75	1.31	13.66	308
RDNOR	1.38	0.79	0.02	0.78	1.24	1.72	5.03	308
RDALT	1.83	1.73	0.01	0.74	1.33	2.2	13.89	308
DIFFP	2.76	2.91	0.06	1.3	1.94	3.24	34.11	308
KS	36.54	20.15	6.59	22.69	31.2	45.55	167.93	308
MIN3CV	1.08	1.01	0.04	0.48	0.81	1.26	8.46	308
MAX3CV	5.49	3.41	0.73	3.14	4.72	6.88	21.17	308
MEAN3CV	3.41	1.94	0.57	2.1	3.04	4.18	15.32	308
MEDIAN3CV	3.46	2	0.63	2.1	3	4.25	16.87	308
MIN3SKEW	-1.48	0.44	-1.73	-1.73	-1.7	-1.42	0.69	308
MAX3SKEW	1.41	0.63	-1.49	1.43	1.68	1.73	1.73	308
MEAN3SKEW	0	0.49	-1.61	-0.26	0.05	0.33	1.16	308
MEDIAN3SKEW	0.03	0.69	-1.7	-0.41	0.07	0.55	1.53	308
MIN3D	2.67	6.13	0.00	0.38	1.03	2.99	73.33	308
MAX3D	26.81	27.14	0.59	10.58	20.08	32.20	200.26	308
MEAN3D	11.62	13.40	0.43	4.63	8.44	13.24	118.06	308
MEDIAN3D	10.88	13.28	0.42	3.89	7.87	12.01	121.30	308
MIN3RD	0.38	0.73	0	0.04	0.14	0.38	6.13	308
MAX3RD	117.7	665.15	0.87	5.26	16.67	54.56	9157.71	308
MEAN3RD	7.53	23.86	0.55	1.92	3.04	5.7	291.4	308
MEDIAN3RD	1.95	2.79	0.29	0.97	1.37	1.98	39.01	308
MIN3RDNOR	0.3	0.33	0	0.06	0.17	0.42	1.63	308
MAX3RDNOR	1.73	0.28	0.76	1.58	1.85	1.95	2	308
MEAN3RDNOR	1	0.22	0.51	0.86	0.98	1.11	1.8	308
MEDIAN3RDNOR	0.99	0.26	0.33	0.81	0.98	1.14	1.85	308
MIN3RDALT	0.27	0.52	0	0.03	0.1	0.27	4.34	308
MAX3RDALT	83.23	470.33	0.62	3.72	11.79	38.58	6475.48	308
MEAN3RDALT	5.32	16.87	0.39	1.36	2.15	4.03	206.05	308
MEDIAN3RDALT	1.38	1.97	0.2	0.68	0.97	1.4	27.58	308
MIN3DIFFP	0.69	0.85	0	0.15	0.44	0.91	8.62	308
MAX3DIFFP	8.1	6.07	1.13	4.26	6.76	10.05	47.45	308
MEAN3DIFFP	3.37	2.32	0.77	1.99	2.77	4.08	22.76	308
MEDIAN3DIFFP	3.13	2.31	0.69	1.79	2.51	3.95	23.91	308

Table 12 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MIN3SPD	0.02	0.02	0	0.01	0.02	0.03	0.18	308
MAX3SPD	0.11	0.07	0.01	0.06	0.09	0.13	0.5	308
MEAN3SPD	0.07	0.04	0.01	0.04	0.06	0.08	0.37	308
MEDIAN3SPD	0.07	0.04	0.01	0.04	0.06	0.08	0.41	308
MIN3KS	26.02	17.06	4.76	14.83	21.48	31.93	137.18	308
MAX3KS	181.77	213.13	12.09	79.31	124.09	206.54	2751.23	308
MEAN3KS	48.26	25.64	7.44	30.53	42.03	60.32	199.62	308
MEDIAN3KS	38.19	20.28	6.19	24.04	33.58	47.94	158.95	308
MIN4CV	1.93	1.66	0.22	0.91	1.5	2.37	15.73	308
MAX4CV	4.74	2.95	0.6	2.75	4.07	5.83	19.16	308
MEAN4CV	3.54	2.01	0.6	2.18	3.13	4.35	15.73	308
MEDIAN4CV	3.6	2.07	0.6	2.22	3.21	4.39	15.73	308
MIN4SKEW	-1.07	0.98	-2	-1.89	-1.5	-0.41	1.78	308
MAX4SKEW	1.07	1.02	-1.93	0.46	1.48	1.87	2	308
MEAN4SKEW	0.04	0.68	-1.93	-0.35	0.07	0.5	1.78	308
MEDIAN4SKEW	0.05	0.71	-1.93	-0.3	0.03	0.54	1.78	308
MIN4D	4.47	11.64	0.00	0.50	1.36	3.86	121.30	308
MAX4D	20.38	20.98	0.36	7.07	14.66	25.33	121.30	308
MEAN4D	10.11	12.76	0.36	3.81	6.99	11.59	121.30	308
MEDIAN4D	9.73	13.27	0.17	3.48	6.40	10.73	121.30	308
MIN4RD	0.68	1.38	0	0.06	0.23	0.67	13.66	308
MAX4RD	9.14	19.83	0.15	1.51	4.31	10.73	298.7	308
MEAN4RD	1.66	1.53	0.15	0.87	1.26	1.94	13.66	308
MEDIAN4RD	1.26	1.37	0.07	0.61	0.87	1.35	13.66	308
MIN4RDNOR	0.53	0.58	0	0.1	0.31	0.77	2.62	308
MAX4RDNOR	1.9	0.72	0.22	1.31	2.09	2.55	2.98	308
MEAN4RDNOR	1.07	0.4	0.22	0.8	1.02	1.27	2.62	308
MEDIAN4RDNOR	1.02	0.44	0.1	0.72	0.93	1.23	2.62	308
MIN4RDALT	0.71	1.44	0	0.07	0.23	0.69	13.89	308
MAX4RDALT	9.67	22.08	0.15	1.55	4.57	11.2	340.47	308
MEAN4RDALT	1.74	1.62	0.15	0.91	1.33	2.02	13.89	308
MEDIAN4RDALT	1.31	1.42	0.07	0.64	0.9	1.38	13.89	308
MIN4DIFFP	1.19	1.93	0	0.19	0.67	1.55	23.91	308
MAX4DIFFP	6.2	5.28	0.64	2.59	4.98	7.89	39.3	308
MEAN4DIFFP	2.95	2.42	0.61	1.66	2.29	3.56	24.48	308
MEDIAN4DIFFP	2.82	2.74	0.2	1.49	2.08	3.28	34.11	308
MIN4KURTO	-2.96	2.98	-6	-5.63	-4.23	-0.28	3.75	308
MAX4KURTO	2.44	1.83	-5.4	1.66	3.14	3.77	4	308
MEAN4KURTO	0.14	1.36	-5.4	-0.58	0.09	0.95	3.75	308
MEDIAN4KURTO	0.34	1.54	-5.4	-0.28	0.61	1.31	3.75	308
MIN4SPD	0.05	0.04	0.01	0.02	0.03	0.05	0.47	308

Table 12 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MAX4SPD	0.11	0.07	0.01	0.06	0.09	0.13	0.5	308
MEAN4SPD	0.08	0.05	0.01	0.05	0.07	0.1	0.47	308
MEDIAN4SPD	0.09	0.05	0.01	0.05	0.08	0.1	0.47	308
MIN4KS	30.68	20.89	5.42	17.65	24.8	36.76	167.93	308
MAX4KS	88.22	69.07	6.59	42.47	67.21	110.66	458.02	308
MEAN4KS	39.97	20.41	6.59	26.34	35.08	50.1	167.93	308
MEDIAN4KS	36.66	19.93	6.59	23.01	31.38	45.53	167.93	308

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

Table 13 Descriptive statistics for the competitive tenders in the Swiss data

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	5.73	1.86	4	4	5	7	13	1082
MEANBIDS	828.06	1803.84	13.63	203.87	423.97	858.67	37786.87	1082
STDBIDS	87.35	216.52	0.41	13.96	32.74	81.27	3996.24	1082
CV	10.12	7.89	0.76	5.91	8.45	11.64	128	1082
KURTO	0.25	2.27	-6	-1.27	0.13	1.75	8.03	1082
SKEW	0.26	0.97	-2.68	-0.37	0.28	0.92	2.47	1082
SPD	2.5	29.79	0.02	0.16	0.24	0.35	730.71	1082
D	54.94	223.18	0.00	5.18	14.19	39.12	4656.85	1082
RD	1.16	2.45	0	0.23	0.57	1.13	41.26	1082
RDNOR	1.04	0.82	0	0.43	0.87	1.44	6.95	1082
RDALT	1.61	3.19	0	0.36	0.84	1.66	47.49	1082
DIFFP	176.79	2246.36	0	1.84	4.36	8.43	50228.95	1082
KS	15.07	10.98	1.48	9.12	12.24	17.44	132.33	1082
MIN3CV	3.37	3.13	0.02	1.16	2.33	4.68	24.05	1082
MAX3CV	14.22	11.46	0.93	8.19	11.77	16.52	122.06	1082
MEAN3CV	9.37	6.9	0.73	5.56	7.92	10.9	91.8	1082
MEDIAN3CV	9.7	7.83	0.6	5.61	8.19	11.41	121.69	1082
MIN3SKEW	-1.38	0.62	-1.73	-1.73	-1.67	-1.34	1.68	1082
MAX3SKEW	1.47	0.53	-1.61	1.48	1.69	1.73	1.73	1082

Table 13 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MEAN3SKEW	0.13	0.58	-1.66	-0.22	0.16	0.51	1.71	1082
MEDIAN3SKEW	0.22	0.83	-1.73	-0.35	0.26	0.83	1.73	1082
MIN3D	13.03	30.15	0.00	1.10	3.79	11.64	364.58	1082
MAX3D	154.93	433.88	0.72	22.68	52.94	132.40	7506.55	1082
MEAN3D	69.05	178.70	0.44	11.05	25.66	60.66	2536.05	1082
MEDIAN3D	64.40	171.21	0.36	10.29	23.09	56.37	2640.97	1082
MIN3RD	0.32	0.68	0	0.03	0.12	0.34	8.83	1082
MAX3RD	69.28	265.38	0.13	4.44	12.27	38.7	5315.77	1082
MEAN3RD	8.2	25.53	0.08	1.55	2.78	6.05	421.57	1082
MEDIAN3RD	2.22	4.95	0.01	0.78	1.21	1.98	72.77	1082
MIN3RDNOR	0.27	0.3	0	0.05	0.16	0.39	1.72	1082
MAX3RDNOR	1.68	0.34	0.17	1.52	1.8	1.93	2	1082
MEAN3RDNOR	0.94	0.27	0.11	0.77	0.93	1.11	1.85	1082
MEDIAN3RDNOR	0.92	0.34	0.01	0.7	0.91	1.12	1.94	1082
MIN3RDALT	0.22	0.48	0	0.02	0.09	0.24	6.24	1082
MAX3RDALT	48.99	187.65	0.1	3.14	8.67	27.36	3758.82	1082
MEAN3RDALT	5.8	18.05	0.06	1.1	1.96	4.27	298.09	1082
MEDIAN3RDALT	1.57	3.5	0	0.55	0.85	1.4	51.46	1082
MIN3DIFFP	2.07	2.77	0	0.36	0.98	2.56	22.67	1082
MAX3DIFFP	226.26	2740.18	0.89	9.94	15.42	23.4	63802.49	1082
MEAN3DIFFP	92.86	1118.61	0.64	4.6	7.07	10.73	23734.93	1082
MEDIAN3DIFFP	60.05	1011.71	0.54	4.08	6.52	10.25	25112.51	1082
MIN3SPD	0.07	0.07	0	0.02	0.05	0.09	0.58	1082
MAX3SPD	2.5	29.79	0.02	0.16	0.24	0.35	730.71	1082
MEAN3SPD	1.14	12.52	0.01	0.11	0.16	0.23	279.23	1082
MEDIAN3SPD	0.76	10.69	0.01	0.11	0.17	0.25	264.32	1082
MIN3KS	10.96	8.72	1.1	6.33	8.72	12.45	107.96	1082
MAX3KS	82.65	179.47	4.21	21.54	43	86.79	4045.44	1082
MEAN3KS	21.45	21.35	2.62	11.62	16.51	24.77	463.06	1082
MEDIAN3KS	15.78	13.43	1.2	9.07	12.57	18.27	167.49	1082
MIN4CV	6.24	6.77	0.07	2.29	4.49	8.14	128	1082
MAX4CV	12.14	9.73	0.76	7.07	10.13	14.17	128	1082
MEAN4CVB	9.82	7.64	0.76	5.74	8.29	11.45	128	1082
MEDIAN4CV	10.24	8.21	0.76	6	8.55	11.74	128	1082
MIN4SKEW	-0.75	1.09	-2	-1.72	-1.02	0	2	1082
MAX4SKEW	1.02	1	-1.99	0.34	1.4	1.85	2	1082
MEAN4SKEW	0.19	0.8	-1.99	-0.31	0.22	0.74	2	1082
MEDIAN4SKEW	0.22	0.86	-1.99	-0.27	0.19	0.79	2	1082
MIN4D	22.95	76.73	0.00	1.76	5.64	18.51	1226.06	1082
MAX4D	111.58	328.01	0.02	13.13	34.32	86.30	5764.43	1082
MEAN4D	59.42	181.05	0.02	8.14	19.23	46.70	3035.79	1082

Table 13 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MEDIAN4D	59.94	224.26	0.00	6.98	18.35	45.41	4656.85	1082
MIN4RD	0.71	2.05	0	0.06	0.2	0.62	41.26	1082
MAX4RD	6.73	17.51	0	0.88	2.55	6.3	266.72	1082
MEAN4RD	1.69	3.23	0	0.56	1	1.72	67.61	1082
MEDIAN4RD	1.24	2.28	0	0.39	0.74	1.25	41.26	1082
MIN4RDNOR	0.51	0.56	0	0.09	0.29	0.76	2.87	1082
MAX4RDNOR	1.62	0.81	0	0.96	1.74	2.3	2.98	1082
MEAN4RDNOR	0.95	0.51	0	0.61	0.9	1.21	2.87	1082
MEDIAN4RDNOR	0.9	0.55	0	0.51	0.84	1.19	2.87	1082
MIN4RDALT	0.75	2.18	0	0.06	0.21	0.67	43.5	1082
MAX4RDALT	7.07	18.5	0	0.93	2.69	6.57	267.55	1082
MEAN4RDALT	1.77	3.36	0	0.59	1.05	1.8	68.45	1082
MEDIAN4RDALT	1.3	2.4	0	0.41	0.78	1.31	43.5	1082
MIN4DIFFP	20.35	549.37	0	0.51	1.54	4.44	18073.22	1082
MAX4DIFFP	206.69	2603.79	0.03	5.44	9.88	16.43	62175.31	1082
MEAN4DIFFP	116.56	1450.99	0.03	3.24	5.46	9.08	30433.18	1082
MEDIAN4DIFFP	140.96	2011.38	0	2.73	5.24	8.95	50228.95	1082
MIN4KURTO	-2.61	3.02	-6	-5.39	-3.37	0.07	4	1082
MAX4KURTO	2.07	2.28	-6	1.43	2.94	3.72	4	1082
MEAN4KURTO	0.12	1.83	-6	-0.83	0.1	1.31	4	1082
MEDIAN4KURTO	0.29	2.06	-6	-0.75	0.6	1.62	4	1082
MIN4SPD	0.34	6.03	0	0.05	0.1	0.2	198.43	1082
MAX4SPD	2.5	29.79	0.02	0.16	0.24	0.35	730.71	1082
MEAN4SPD	1.54	17.19	0.02	0.14	0.2	0.29	388.02	1082
MEDIAN4SPD	1.78	23.58	0.02	0.14	0.21	0.3	621.75	1082
MIN4KS	12.89	10.47	1.38	7.42	10.24	14.5	132.33	1082
MAX4KS	36.64	66.05	1.48	12.61	22.53	44.04	1433.11	1082
MEAN4KS	16.65	11.9	1.48	9.73	13.55	19.78	132.33	1082
MEDIAN4KS	14.87	11.3	1.48	8.9	12.14	16.99	132.33	1082

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

Table 14 Descriptive statistics for incomplete bid-rigging cartels in sample 1 (Swiss data)

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	7.49	2.54	4	5	7	9	13	252
MEANBIDS	435.90	476.67	18.01	146.59	296.62	538.73	3460.91	252
STDBIDS	35.12	50.17	1.54	9.60	19.62	33.72	362.86	252
CV	7.79	3.89	1.77	5.14	6.79	9.6	23.92	252
KURTO	0.41	2.12	-5.9	-0.94	0.04	1.54	6.97	252
SKEW	-0.07	0.99	-2.59	-0.67	-0.06	0.5	2.57	252
SPD	0.26	0.16	0.05	0.16	0.21	0.32	0.89	252
D	21.07	39.03	0.08	4.33	9.10	20.87	351.83	252
RD	1.39	2.67	0.01	0.26	0.58	1.42	28.37	252
RDNOR	1.41	1.11	0.01	0.59	1.14	2	5.48	252
RDALT	2.32	3.52	0.01	0.55	1.17	2.65	28.54	252
DIFFP	6.34	8.85	0.03	1.8	3.69	7.02	73.53	252
KS	16.62	8.04	4.15	10.99	15.24	19.97	57.54	252
MIN3CV	1.68	1.73	0	0.65	1.24	2.14	14.77	252
MAX3CV	12.26	6.39	2.9	7.87	10.66	15.21	38.48	252
MEAN3CV	7.12	3.48	1.55	4.7	6.32	8.98	22.96	252
MEDIAN3CV	7.06	3.81	1.01	4.59	6.28	8.51	29.82	252
MIN3SKEW	-1.6	0.36	-1.73	-1.73	-1.72	-1.65	0.97	252
MAX3SKEW	1.53	0.48	-1.1	1.64	1.72	1.73	1.73	252
MEAN3SKEW	-0.04	0.5	-1.43	-0.3	-0.03	0.25	1.47	252
MEDIAN3SKEW	-0.05	0.77	-1.71	-0.59	-0.05	0.51	1.73	252
MIN3D	3.64	8.58	0.00	0.43	1.31	3.23	80.15	252
MAX3D	76.68	107.56	1.58	18.87	41.15	80.60	842.70	252
MEAN3D	29.56	44.27	0.62	8.00	16.57	28.64	437.36	252
MEDIAN3D	26.54	40.57	0.41	7.70	14.60	27.21	424.30	252
MIN3RD	0.23	0.45	0	0.02	0.07	0.19	3.35	252
MAX3RD	1203.97	9993.9	0.68	10.98	31.05	96.88	122393.82	252
MEAN3RD	27.54	191.86	0.31	2.5	4.5	9.16	2742.11	252
MEDIAN3RD	2.09	2.42	0.21	1.02	1.49	2.21	27.08	252
MIN3RDNOR	0.21	0.28	0	0.03	0.09	0.24	1.41	252
MAX3RDNOR	1.83	0.23	0.65	1.77	1.92	1.97	2	252
MEAN3RDNOR	1.02	0.22	0.33	0.89	1.02	1.12	1.73	252
MEDIAN3RDNOR	1.02	0.3	0.26	0.83	1.02	1.2	1.89	252
MIN3RDALT	0.16	0.32	0	0.02	0.05	0.14	2.37	252
MAX3RDALT	851.33	7066.76	0.48	7.76	21.95	68.51	86545.5	252
MEAN3RDALT	19.47	135.66	0.22	1.77	3.18	6.48	1938.96	252
MEDIAN3RDALT	1.48	1.71	0.15	0.72	1.06	1.56	19.15	252
MIN3DIFFP	0.97	1.58	0	0.17	0.47	1.15	12.6	252
MAX3DIFFP	19.61	12.81	2.29	11.28	16.77	25.1	87.65	252
MEAN3DIFFP	7.76	5.62	1.39	4.79	6.4	8.94	49.6	252
MEDIAN3DIFFP	7.11	6.59	1.14	3.9	5.72	8.21	73.53	252

Table 14 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MIN3SPD	0.03	0.04	0	0.01	0.02	0.04	0.32	252
MAX3SPD	0.26	0.16	0.05	0.16	0.21	0.32	0.89	252
MEAN3SPD	0.15	0.08	0.03	0.09	0.13	0.19	0.62	252
MEDIAN3SPD	0.15	0.09	0.02	0.09	0.13	0.18	0.83	252
MIN3KS	10.96	6.18	3	6.88	9.62	12.9	34.52	252
MAX3KS	357.52	3429.85	6.79	47.14	80.58	155	54476.99	252
MEAN3KS	27.25	34.52	6.03	15.6	21.55	29.85	495.16	252
MEDIAN3KS	18.68	11.16	3.36	12.03	16.1	22.21	99.02	252
MIN4CV	3.16	2.97	0.12	1.31	2.26	4.2	23.92	252
MAX4CV	10.63	5.46	2.38	6.97	9.43	13.25	33.94	252
MEAN4CV	7.43	3.65	1.64	4.89	6.54	9.33	23.92	252
MEDIAN4CV	7.65	3.91	1.69	5.02	6.76	9.66	25.38	252
MIN4SKEW	-1.42	0.79	-2	-1.95	-1.78	-1.22	1.87	252
MAX4SKEW	1.28	0.97	-1.98	1.09	1.79	1.94	2	252
MEAN4SKEW	-0.07	0.69	-1.98	-0.43	-0.05	0.38	1.87	252
MEDIAN4SKEW	-0.07	0.79	-1.98	-0.47	-0.03	0.32	1.87	252
MIN4D	4.98	11.53	0.00	0.48	1.63	4.68	102.18	252
MAX4D	60.49	87.82	0.18	13.66	33.62	66.93	771.65	252
MEAN4D	25.08	38.97	0.18	6.81	14.32	25.80	410.17	252
MEDIAN4D	24.93	42.34	0.14	6.02	13.17	25.38	424.30	252
MIN4RD	0.59	2.26	0	0.03	0.09	0.31	28.37	252
MAX4RD	25.25	166.8	0.04	3.1	7.14	16.87	2627.72	252
MEAN4RD	2.25	2.96	0.04	0.91	1.52	2.37	28.37	252
MEDIAN4RD	1.53	2.43	0.04	0.53	1.01	1.66	28.37	252
MIN4RDNOR	0.37	0.54	0	0.05	0.15	0.42	2.8	252
MAX4RDNOR	2.18	0.68	0.07	1.83	2.37	2.7	3	252
MEAN4RDNOR	1.11	0.46	0.07	0.8	1.08	1.33	2.8	252
MEDIAN4RDNOR	1.06	0.54	0.07	0.66	1.05	1.38	2.8	252
MIN4RDALT	0.61	2.29	0	0.03	0.1	0.32	28.54	252
MAX4RDALT	27.16	186.06	0.04	3.13	7.47	18.11	2933.56	252
MEAN4RDALT	2.35	3.08	0.04	0.96	1.62	2.48	28.54	252
MEDIAN4RDALT	1.59	2.46	0.04	0.56	1.08	1.7	28.54	252
MIN4DIFFP	1.83	4.55	0	0.19	0.58	1.58	57.15	252
MAX4DIFFP	16.09	11.96	0.3	8.81	13.53	20.12	77.43	252
MEAN4DIFFP	7.05	6.61	0.3	3.75	5.57	7.76	60.05	252
MEDIAN4DIFFP	6.8	7.49	0.24	3.19	5.09	7.32	73.53	252
MIN4KURTO	-4	2.64	-6	-5.86	-5.28	-3.15	3.94	252
MAX4KURTO	2.99	1.68	-5.9	2.79	3.68	3.91	4	252
MEAN4KURTO	0.16	1.38	-5.9	-0.48	0.04	0.84	3.94	252
MEDIAN4KURTO	0.51	1.67	-5.9	-0.08	0.64	1.49	3.94	252
MIN4SPRD	0.08	0.08	0	0.03	0.05	0.09	0.81	252

Table 14 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MAX4SPD	0.26	0.16	0.05	0.16	0.21	0.32	0.89	252
MEAN4SPD	0.19	0.11	0.04	0.12	0.16	0.23	0.81	252
MEDIAN4SPD	0.19	0.12	0.04	0.12	0.16	0.24	0.89	252
MIN4KS	12.65	7.28	3.64	7.85	10.91	14.56	42.15	252
MAX4KS	64.97	79.07	4.15	24.21	44.41	76.85	862.88	252
MEAN4KS	19.96	11.22	4.15	12.92	17.51	23.75	97.43	252
MEDIAN4KS	16.85	8.23	3.87	11.12	15.26	20.29	59.16	252

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

Table 15 Descriptive statistics for incomplete bid-rigging cartels in sample 2 (Swiss data)

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	7.77	2.47	4	6	8	10	13	223
MEANBIDS	405.24	372.35	27.84	149.78	302.21	535.40	3002.37	223
STDBIDS	30.96	37.64	1.63	9.65	19.51	33.19	270.82	223
CV	7.6	3.77	1.77	5.14	6.66	9.07	23.92	223
KURTO	0.44	2.08	-5.75	-0.94	0.04	1.47	6.97	223
SKEW	-0.07	0.97	-2.59	-0.62	-0.08	0.46	2.57	223
SPD	0.26	0.16	0.05	0.16	0.21	0.31	0.89	223
D	20.13	36.04	0.08	4.13	8.90	21.08	351.83	223
RD	1.29	2.12	0.01	0.27	0.58	1.36	18.04	223
RDNOR	1.45	1.14	0.01	0.61	1.14	2.06	5.48	223
RDALT	2.3	3.24	0.01	0.58	1.17	2.83	20.74	223
DIFFP	6.3	9.13	0.03	1.88	3.69	6.78	73.53	223
KS	16.85	7.91	4.15	11.89	15.54	19.99	57.54	223
MIN3CV	1.47	1.22	0	0.62	1.15	1.93	7.05	223
MAX3CV	12.24	6.43	2.93	8	10.57	14.98	38.48	223
MEAN3CV	6.92	3.33	1.55	4.69	6.09	8.33	22.96	223
MEDIAN3CV	6.75	3.56	1.01	4.58	6	8.06	29.82	223
MIN3SKEW	-1.63	0.29	-1.73	-1.73	-1.72	-1.68	0.62	223
MAX3SKEW	1.58	0.38	-1.1	1.68	1.72	1.73	1.73	223
MEAN3SKEW	-0.04	0.45	-1.33	-0.28	-0.04	0.25	1.31	223

Table 15 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MEDIAN3SKEW	-0.05	0.74	-1.69	-0.58	-0.06	0.45	1.73	223
MIN3D	2.76	5.42	0.00	0.38	1.17	2.87	56.23	223
MAX3D	69.79	84.97	2.38	20.13	42.35	79.53	670.57	223
MEAN3D	26.13	30.47	1.42	8.28	16.72	28.29	235.67	223
MEDIAN3D	23.11	25.43	1.52	7.92	14.63	26.60	171.87	223
MIN3RD	0.18	0.37	0	0.02	0.06	0.15	3.35	223
MAX3RD	1355.45	10617.14	0.92	14.74	37.36	128.17	122393.82	223
MEAN3RD	30.39	203.82	0.43	2.65	4.66	9.35	2742.11	223
MEDIAN3RD	1.92	1.72	0.26	1.06	1.5	2.14	18.63	223
MIN3RDNOR	0.18	0.25	0	0.03	0.08	0.19	1.41	223
MAX3RDNOR	1.85	0.2	0.79	1.82	1.93	1.98	2	223
MEAN3RDNOR	1.02	0.2	0.41	0.9	1.02	1.12	1.66	223
MEDIAN3RDNOR	1.02	0.28	0.3	0.85	1.02	1.2	1.85	223
MIN3RDALT	0.13	0.26	0	0.01	0.04	0.11	2.37	223
MAX3RDALT	958.45	7507.45	0.65	10.42	26.42	90.63	86545.5	223
MEAN3RDALT	21.49	144.12	0.3	1.88	3.29	6.61	1938.96	223
MEDIAN3RDALT	1.36	1.21	0.18	0.75	1.06	1.51	13.18	223
MIN3DIFFP	0.81	1.25	0	0.14	0.43	0.99	10.47	223
MAX3DIFFP	19.73	12.93	2.29	11.81	16.81	24.42	87.65	223
MEAN3DIFFP	7.59	5.58	1.39	4.82	6.35	8.75	49.6	223
MEDIAN3DIFFP	6.92	6.67	1.14	3.91	5.71	8.1	73.53	223
MIN3SPD	0.03	0.02	0	0.01	0.02	0.04	0.15	223
MAX3SPD	0.26	0.16	0.05	0.16	0.21	0.31	0.89	223
MEAN3SPD	0.15	0.08	0.03	0.09	0.13	0.18	0.62	223
MEDIAN3SPD	0.14	0.09	0.02	0.09	0.12	0.17	0.83	223
MIN3KS	10.9	6.01	3	6.98	9.7	12.52	34.52	223
MAX3KS	394.81	3645.21	14.27	52.14	87.43	161.06	54476.99	223
MEAN3KS	27.97	36.27	6.25	16	21.99	30.38	495.16	223
MEDIAN3KS	19.22	11.35	3.36	12.64	16.97	22.22	99.02	223
MIN4CV	2.8	2.48	0.12	1.2	2.17	3.87	23.92	223
MAX4CV	10.63	5.48	2.48	7.05	9.36	13.08	33.94	223
MEAN4CV	7.22	3.5	1.64	4.88	6.32	8.58	23.92	223
MEDIAN4CV	7.41	3.75	1.69	5.01	6.51	8.9	25.38	223
MIN4SKEW	-1.51	0.71	-2	-1.95	-1.82	-1.42	1.84	223
MAX4SKEW	1.38	0.85	-1.96	1.23	1.82	1.94	2	223
MEAN4SKEW	-0.06	0.64	-1.96	-0.4	-0.07	0.35	1.84	223
MEDIAN4SKEW	-0.07	0.74	-1.96	-0.46	-0.04	0.29	1.84	223
MIN4D	3.82	8.45	0.00	0.43	1.43	3.77	96.23	223
MAX4D	58.39	72.08	0.18	16.06	37.52	68.97	535.26	223
MEAN4D	23.35	29.62	0.18	7.26	14.45	25.83	274.73	223
MEDIAN4D	23.01	32.93	0.14	6.07	13.49	25.40	351.83	223

Table 15 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MIN4RD	0.41	1.4	0	0.03	0.09	0.25	18.04	223
MAX4RD	27.52	177.13	0.04	3.91	7.86	17.75	2627.72	223
MEAN4RD	2.19	2.53	0.04	0.96	1.62	2.37	24.7	223
MEDIAN4RD	1.42	1.77	0.04	0.57	1.02	1.52	18.04	223
MIN4RDNOR	0.32	0.46	0	0.04	0.14	0.36	2.7	223
MAX4RDNOR	2.25	0.61	0.07	2	2.42	2.71	3	223
MEAN4RDNOR	1.11	0.42	0.07	0.82	1.09	1.3	2.7	223
MEDIAN4RDNOR	1.06	0.51	0.07	0.69	1.06	1.31	2.7	223
MIN4RDALT	0.42	1.42	0	0.03	0.09	0.27	18.05	223
MAX4RDALT	29.63	197.6	0.04	4	8.35	18.57	2933.56	223
MEAN4RDALT	2.29	2.67	0.04	1.02	1.72	2.49	27.4	223
MEDIAN4RDALT	1.48	1.8	0.04	0.6	1.1	1.55	18.05	223
MIN4DIFFP	1.48	4.34	0	0.17	0.51	1.23	57.15	223
MAX4DIFFP	16.59	11.97	0.3	9.6	13.87	20.28	77.43	223
MEAN4DIFFP	7	6.68	0.3	3.97	5.51	7.55	60.05	223
MEDIAN4DIFFP	6.74	7.66	0.24	3.24	5.08	7.23	73.53	223
MIN4KURTO	-4.32	2.38	-6	-5.88	-5.48	-3.92	3.85	223
MAX4KURTO	3.16	1.43	-5.75	3.04	3.74	3.91	4	223
MEAN4KURTO	0.15	1.23	-5.75	-0.48	0.04	0.79	3.85	223
MEDIAN4KURTO	0.52	1.56	-5.75	-0.05	0.63	1.42	3.85	223
MIN4SPD	0.07	0.07	0	0.03	0.05	0.09	0.81	223
MAX4SPD	0.26	0.16	0.05	0.16	0.21	0.31	0.89	223
MEAN4SPD	0.18	0.11	0.04	0.12	0.15	0.22	0.81	223
MEDIAN4SPD	0.19	0.11	0.04	0.12	0.16	0.22	0.89	223
MIN4KS	12.55	7.02	3.64	7.95	11.06	14.33	40.78	223
MAX4KS	69.75	82.58	4.15	26.17	46.65	84.1	862.88	223
MEAN4KS	20.49	11.41	4.15	13.41	18.45	23.97	97.43	223
MEDIAN4KS	17.14	8.1	3.87	11.75	15.89	20.31	59.16	223

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

Table 16 Descriptive statistics for incomplete bid-rigging cartels in sample 3 (Swiss data)

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	8.5	2.15	5	7	8	10	13	173
MEANBIDS	439.36	388.89	30.24	182.64	329.61	572.01	3002.37	173
STDBIDS	33.50	40.38	2.01	11.89	20.48	33.44	270.82	173
CV	7.54	3.26	1.77	5.24	6.76	9.12	21.7	173
KURTO	0.46	1.97	-2.69	-0.88	-0.09	1.23	6.97	173
SKEW	-0.07	0.95	-2.59	-0.56	-0.08	0.43	2.57	173
SPD	0.26	0.14	0.06	0.17	0.22	0.31	0.88	173
D	21.95	39.01	0.08	4.53	9.46	22.94	351.83	173
RD	1.14	1.76	0.01	0.29	0.58	1.19	12.47	173
RDNOR	1.54	1.17	0.02	0.66	1.17	2.17	5.48	173
RDALT	2.33	3.2	0.02	0.63	1.19	2.81	20.74	173
DIFFP	5.91	7.48	0.1	2.13	3.7	6.74	45.42	173
KS	16.55	7.64	6.19	11.84	15.48	19.69	57.54	173
MIN3CV	1.2	0.9	0	0.52	1.03	1.56	3.95	173
MAX3CV	12.6	5.86	2.93	8.78	10.99	15.19	38.48	173
MEAN3CV	6.82	2.83	1.55	4.79	6.11	8.34	16.5	173
MEDIAN3CV	6.51	2.79	1.01	4.67	5.92	8.02	14.2	173
MIN3SKEW	-1.69	0.12	-1.73	-1.73	-1.73	-1.71	-0.88	173
MAX3SKEW	1.67	0.15	0.73	1.69	1.73	1.73	1.73	173
MEAN3SKEW	-0.03	0.39	-1.02	-0.22	-0.04	0.23	0.88	173
MEDIAN3SKEW	-0.03	0.69	-1.48	-0.47	-0.06	0.41	1.73	173
MIN3D	2.35	4.05	0.00	0.34	1.03	2.40	30.19	173
MAX3D	78.60	90.43	3.70	24.82	51.54	89.92	670.57	173
MEAN3D	28.43	32.44	1.58	9.40	17.96	32.97	235.67	173
MEDIAN3D	24.31	26.46	1.52	8.53	15.79	26.36	171.87	173
MIN3RD	0.11	0.19	0	0.02	0.04	0.12	1.76	173
MAX3RD	1737.96	12034.74	2.68	20.39	48.9	163.12	122393.82	173
MEAN3RD	37.63	231.03	0.92	2.91	5.15	9.8	2742.11	173
MEDIAN3RD	1.72	1.05	0.32	1.07	1.48	1.95	6.05	173
MIN3RDNOR	0.12	0.16	0	0.02	0.06	0.16	1	173
MAX3RDNOR	1.9	0.13	1.31	1.87	1.94	1.98	2	173
MEAN3RDNOR	1.02	0.18	0.61	0.91	1.03	1.11	1.46	173
MEDIAN3RDNOR	1.02	0.25	0.37	0.86	1.02	1.16	1.62	173
MIN3RDALT	0.08	0.14	0	0.01	0.03	0.09	1.25	173
MAX3RDALT	1228.92	8509.85	1.9	14.42	34.57	115.34	86545.5	173
MEAN3RDALT	26.61	163.36	0.65	2.06	3.64	6.93	1938.96	173
MEDIAN3RDALT	1.22	0.74	0.22	0.76	1.05	1.38	4.28	173
MIN3DIFFP	0.58	0.71	0	0.12	0.37	0.81	4.49	173
MAX3DIFFP	20.62	11.43	3.84	13.06	17.99	25.01	72.01	173
MEAN3DIFFP	7.33	3.89	1.55	5	6.42	8.68	22.8	173
MEDIAN3DIFFP	6.18	3.26	1.33	3.91	5.67	7.45	21.76	173

Table 16 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MIN3SPD	0.02	0.02	0	0.01	0.02	0.03	0.08	173
MAX3SPD	0.26	0.14	0.06	0.17	0.22	0.31	0.88	173
MEAN3SPD	0.14	0.07	0.03	0.1	0.13	0.18	0.38	173
MEDIAN3SPD	0.13	0.06	0.02	0.09	0.12	0.17	0.3	173
MIN3KS	10.03	5.06	3	6.91	9.18	11.63	34.52	173
MAX3KS	489.07	4136.34	25.56	63.95	97.45	192.63	54476.99	173
MEAN3KS	29.3	40.7	9.3	16.34	21.99	30.15	495.16	173
MEDIAN3KS	19.58	12.05	7.28	12.7	17.13	21.76	99.02	173
MIN4CV	2.14	1.45	0.12	1.03	1.76	2.94	8.25	173
MAX4CV	11.02	5.04	2.48	7.92	9.6	13.24	33.94	173
MEAN4CV	7.12	2.98	1.64	4.99	6.32	8.58	18.01	173
MEDIAN4CV	7.27	3.09	1.69	5.15	6.57	8.91	17.17	173
MIN4SKEW	-1.69	0.44	-2	-1.96	-1.88	-1.64	-0.04	173
MAX4SKEW	1.62	0.52	-0.4	1.49	1.84	1.95	2	173
MEAN4SKEW	-0.05	0.55	-1.42	-0.34	-0.07	0.31	1.32	173
MEDIAN4SKEW	-0.06	0.67	-1.86	-0.42	-0.05	0.27	1.72	173
MIN4D	2.86	4.84	0.00	0.34	1.13	3.02	30.28	173
MAX4D	66.68	76.42	1.56	20.02	43.87	78.43	535.26	173
MEAN4D	25.49	31.63	1.22	8.73	15.85	27.51	274.73	173
MEDIAN4D	24.96	35.20	0.96	7.98	15.12	27.02	351.83	173
MIN4RD	0.16	0.26	0	0.02	0.06	0.17	1.93	173
MAX4RD	33.76	200.71	1.09	5.32	10.34	20.54	2627.72	173
MEAN4RD	2.22	2.44	0.35	1.04	1.68	2.38	24.7	173
MEDIAN4RD	1.31	1.24	0.16	0.69	1.05	1.39	9.81	173
MIN4RDNOR	0.2	0.25	0	0.03	0.1	0.25	1.51	173
MAX4RDNOR	2.42	0.44	1.07	2.2	2.53	2.75	3	173
MEAN4RDNOR	1.11	0.35	0.43	0.85	1.1	1.28	2.06	173
MEDIAN4RDNOR	1.07	0.44	0.22	0.79	1.07	1.25	2.5	173
MIN4RDALT	0.16	0.26	0	0.02	0.07	0.18	2.03	173
MAX4RDALT	36.41	223.95	1.1	5.54	10.76	21.56	2933.56	173
MEAN4RDALT	2.31	2.61	0.37	1.09	1.75	2.51	27.4	173
MEDIAN4RDALT	1.37	1.27	0.16	0.72	1.11	1.43	9.91	173
MIN4DIFFP	0.71	0.9	0	0.13	0.41	0.98	6.52	173
MAX4DIFFP	17.69	10.69	3.03	11.06	15.55	21.32	70.09	173
MEAN4DIFFP	6.67	4.5	1.18	4.22	5.62	7.44	26.42	173
MEDIAN4DIFFP	6.33	5.07	0.98	3.44	5.23	7.15	34.35	173
MIN4KURTO	-4.9	1.6	-6	-5.92	-5.55	-4.6	1.44	173
MAX4KURTO	3.56	0.61	1	3.49	3.82	3.93	4	173
MEAN4KURTO	0.17	0.89	-1.78	-0.46	-0.05	0.63	3.19	173
MEDIAN4KURTO	0.63	1.19	-4.42	0	0.63	1.34	3.57	173
MIN4SPD	0.05	0.03	0	0.02	0.04	0.07	0.2	173

Table 16 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MAX4SPD	0.26	0.14	0.06	0.17	0.22	0.31	0.88	173
MEAN4SPD	0.18	0.08	0.04	0.12	0.15	0.22	0.48	173
MEDIAN4SPD	0.18	0.09	0.04	0.12	0.16	0.22	0.48	173
MIN4KS	11.44	5.86	3.64	7.8	10.66	12.95	40.78	173
MAX4KS	81.53	89.76	12.72	34.28	57.17	97.8	862.88	173
MEAN4KS	20.95	12.1	8.31	13.75	18.47	23.97	97.43	173
MEDIAN4KS	16.96	7.85	5.9	11.68	15.89	19.94	59.16	173

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

Table 17 Descriptive statistics for incomplete bid-rigging cartels in sample 4 (Swiss data)

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	9.08	1.87	6	8	9	11	13	135
MEANBIDS	448.31	384.31	41.46	219.13	337.02	567.20	3002.37	135
STDBIDS	31.74	34.92	2.26	13.09	21.49	32.73	270.82	135
CV	7.19	2.93	1.77	5.22	6.3	8.62	18.9	135
KURTO	0.47	1.97	-2.26	-0.88	-0.18	1.18	6.97	135
SKEW	-0.09	0.94	-2.59	-0.54	-0.12	0.44	2.57	135
SPD	0.26	0.13	0.06	0.17	0.21	0.31	0.76	135
D	22.85	42.70	0.26	4.76	9.34	22.35	351.83	135
RD	1.05	1.5	0.01	0.3	0.57	1.12	10.58	135
RDNOR	1.6	1.18	0.04	0.7	1.22	2.28	5.48	135
RDALT	2.33	3.02	0.04	0.67	1.27	2.86	20.74	135
DIFFP	5.82	7.82	0.1	2.03	3.36	6.58	45.42	135
KS	16.91	7.21	6.62	12.48	16.29	19.83	57.54	135
MIN3CV	0.94	0.62	0	0.43	0.82	1.28	3.04	135
MAX3CV	12.33	5.32	3.36	8.87	10.64	15.02	32.54	135
MEAN3CV	6.47	2.56	1.55	4.69	5.85	7.72	15.38	135
MEDIAN3CV	6.07	2.49	1.01	4.52	5.67	7.29	13.93	135
MIN3SKEW	-1.71	0.07	-1.73	-1.73	-1.73	-1.72	-1.13	135
MAX3SKEW	1.69	0.1	0.93	1.71	1.73	1.73	1.73	135

Table 17 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MEAN3SKEW	-0.04	0.37	-1.02	-0.24	-0.04	0.21	0.87	135
MEDIAN3SKEW	-0.03	0.66	-1.45	-0.47	-0.07	0.4	1.73	135
MIN3D	1.96	3.70	0.00	0.30	0.87	1.99	30.19	135
MAX3D	81.04	93.45	5.67	30.62	55.61	88.77	670.57	135
MEAN3D	28.50	33.46	2.14	11.22	18.00	30.40	235.67	135
MEDIAN3D	23.43	25.63	2.24	9.00	16.39	25.58	171.87	135
MIN3RD	0.09	0.18	0	0.02	0.04	0.1	1.76	135
MAX3RD	2203.8	13597.91	3.46	27.6	62.53	192.17	122393.82	135
MEAN3RD	45.46	261.04	0.92	3.13	5.48	10.57	2742.11	135
MEDIAN3RD	1.69	0.96	0.45	1.08	1.5	1.95	5.61	135
MIN3RDNOR	0.1	0.14	0	0.02	0.06	0.13	1	135
MAX3RDNOR	1.92	0.1	1.42	1.91	1.96	1.99	2	135
MEAN3RDNOR	1.02	0.17	0.61	0.92	1.04	1.11	1.46	135
MEDIAN3RDNOR	1.02	0.24	0.49	0.87	1.02	1.16	1.6	135
MIN3RDALT	0.06	0.13	0	0.01	0.03	0.07	1.25	135
MAX3RDALT	1558.32	9615.17	2.44	19.52	44.22	135.89	86545.5	135
MEAN3RDALT	32.15	184.58	0.65	2.22	3.88	7.48	1938.96	135
MEDIAN3RDALT	1.2	0.68	0.32	0.76	1.06	1.38	3.97	135
MIN3DIFFP	0.44	0.47	0	0.11	0.27	0.56	2.32	135
MAX3DIFFP	20.71	11.13	3.84	13.33	17.81	25.2	63.38	135
MEAN3DIFFP	7.11	3.99	1.55	4.82	6.24	8.15	22.8	135
MEDIAN3DIFFP	5.68	2.93	1.33	3.81	5.29	6.99	21.76	135
MIN3SPD	0.02	0.01	0	0.01	0.02	0.02	0.06	135
MAX3SPD	0.26	0.13	0.06	0.17	0.21	0.31	0.76	135
MEAN3SPD	0.14	0.06	0.03	0.1	0.12	0.16	0.35	135
MEDIAN3SPD	0.12	0.05	0.02	0.09	0.11	0.15	0.3	135
MIN3KS	9.9	4.37	3.47	7	9.69	11.51	30.19	135
MAX3KS	603.24	4679.33	32.93	77.85	122.26	231.34	54477	135
MEAN3KS	30.89	43.76	10.24	18.79	23.63	31.72	495.16	135
MEDIAN3KS	20.34	11.39	7.28	14.06	17.92	22.59	99.02	135
MIN4CV	1.7	1.03	0.12	0.91	1.5	2.18	5.53	135
MAX4CV	10.83	4.53	2.8	8	9.4	13.15	28.01	135
MEAN4CV	6.76	2.69	1.64	4.93	6.08	8	16.54	135
MEDIAN4CV	6.88	2.88	1.69	5.01	6.17	8.11	17.17	135
MIN4SKEW	-1.81	0.29	-2	-1.97	-1.92	-1.76	-0.28	135
MAX4SKEW	1.72	0.4	-0.26	1.7	1.86	1.96	2	135
MEAN4SKEW	-0.06	0.52	-1.37	-0.33	-0.1	0.28	1.18	135
MEDIAN4SKEW	-0.07	0.65	-1.69	-0.38	-0.07	0.26	1.72	135
MIN4D	2.38	4.46	0.00	0.33	1.04	2.48	30.28	135
MAX4D	69.99	79.50	5.47	25.08	47.56	78.76	535.26	135
MEAN4D	25.93	33.56	1.39	9.45	16.57	26.24	274.73	135

Table 17 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MEDIAN4D	25.13	37.53	0.96	7.99	16.21	25.40	351.83	135
MIN4RD	0.12	0.19	0	0.02	0.05	0.14	1.47	135
MAX4RD	41.1	226.79	1.09	6.95	13.21	23.81	2627.72	135
MEAN4RD	2.27	2.46	0.35	1.11	1.72	2.51	24.7	135
MEDIAN4RD	1.25	0.95	0.16	0.69	1.06	1.36	5.95	135
MIN4RDNOR	0.16	0.19	0	0.03	0.09	0.21	1.02	135
MAX4RDNOR	2.53	0.35	1.07	2.34	2.61	2.79	3	135
MEAN4RDNOR	1.11	0.33	0.43	0.87	1.11	1.28	2.05	135
MEDIAN4RDNOR	1.07	0.42	0.22	0.8	1.07	1.25	2.27	135
MIN4RDALT	0.12	0.17	0	0.02	0.06	0.15	1.04	135
MAX4RDALT	44.43	253.08	1.1	7.07	13.34	26.04	2933.56	135
MEAN4RDALT	2.37	2.65	0.37	1.16	1.79	2.62	27.4	135
MEDIAN4RDALT	1.31	1	0.16	0.73	1.11	1.43	6.3	135
MIN4DIFFP	0.53	0.59	0	0.12	0.37	0.81	3.21	135
MAX4DIFFP	18.02	10.4	3.03	11.48	15.55	22.07	56.74	135
MEAN4DIFFP	6.55	4.74	1.18	4.08	5.51	7.01	26.42	135
MEDIAN4DIFFP	6.13	5.32	0.98	3.26	4.82	6.87	34.35	135
MIN4KURTO	-5.26	1.06	-6	-5.93	-5.69	-5.08	-0.16	135
MAX4KURTO	3.71	0.39	2.18	3.62	3.86	3.95	4	135
MEAN4KURTO	0.16	0.8	-1.21	-0.4	-0.07	0.62	2.41	135
MEDIAN4KURTO	0.72	0.98	-3.63	0.06	0.68	1.38	3.11	135
MIN4SPD	0.04	0.02	0	0.02	0.03	0.05	0.13	135
MAX4SPD	0.26	0.13	0.06	0.17	0.21	0.31	0.76	135
MEAN4SPD	0.17	0.08	0.04	0.12	0.15	0.2	0.45	135
MEDIAN4SPD	0.17	0.08	0.04	0.12	0.15	0.2	0.48	135
MIN4KS	11.21	5.03	4.25	7.95	10.99	12.79	36.12	135
MAX4KS	91.93	91.08	18.25	46.29	67.17	110.97	862.88	135
MEAN4KS	21.62	10.72	8.43	15.07	19.14	24.63	97.28	135
MEDIAN4KS	17.62	7.69	5.9	13.03	16.55	20.31	59.16	135

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

Table 18 Descriptive statistics for incomplete bid-rigging cartels in sample 5 (Swiss data)

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
NBRBIDS	9.42	1.81	7	8	9	11	13	104
MEANBIDS	434.09	290.66	86.12	232.84	345.62	534.23	1559.96	104
STDBIDS	29.66	26.44	4.58	15.43	21.89	32.70	191.92	104
CV	6.87	2.53	1.77	5.19	6.22	8.05	14.4	104
KURTO	0.52	2.06	-2.26	-0.89	-0.15	1.17	6.97	104
SKEW	-0.1	0.94	-2.59	-0.62	-0.16	0.43	2.57	104
SPD	0.25	0.12	0.06	0.18	0.21	0.28	0.66	104
D	23.61	46.99	0.26	5.33	8.88	21.49	351.83	104
RD	1.03	1.56	0.01	0.3	0.57	1.09	10.58	104
RDNOR	1.65	1.24	0.04	0.72	1.25	2.39	5.48	104
RDALT	2.4	3.23	0.04	0.69	1.29	3.01	20.74	104
DIFFP	5.91	8.38	0.1	1.77	3.22	6.14	45.42	104
KS	17.25	7.04	6.93	13.12	16.66	19.81	57.54	104
MIN3CV	0.86	0.6	0	0.38	0.75	1.21	3.04	104
MAX3CV	11.97	4.77	3.36	8.96	10.54	13.64	26.24	104
MEAN3CV	6.19	2.19	1.55	4.7	5.74	7.34	12.47	104
MEDIAN3CV	5.71	2.04	1.55	4.39	5.55	6.88	11.38	104
MIN3SKEW	-1.71	0.08	-1.73	-1.73	-1.73	-1.72	-1.13	104
MAX3SKEW	1.71	0.05	1.48	1.71	1.73	1.73	1.73	104
MEAN3SKEW	-0.04	0.35	-1.02	-0.25	-0.05	0.16	0.87	104
MEDIAN3SKEW	-0.05	0.62	-1.45	-0.47	-0.07	0.36	1.32	104
MIN3D	1.47	2.09	0.00	0.31	0.81	1.65	13.14	104
MAX3D	76.95	75.24	9.36	35.44	56.91	86.41	530.84	104
MEAN3D	27.35	30.06	3.76	12.85	18.95	28.64	235.67	104
MEDIAN3D	21.88	21.69	2.24	10.65	16.45	24.90	144.14	104
MIN3RD	0.06	0.07	0	0.01	0.04	0.09	0.34	104
MAX3RD	2673.49	15390.37	3.46	35.67	75.98	246.78	122393.82	104
MEAN3RD	53.44	294.68	0.92	3.42	5.6	10.88	2742.11	104
MEDIAN3RD	1.66	0.88	0.45	1.11	1.48	1.95	5.61	104
MIN3RDNOR	0.08	0.08	0	0.02	0.06	0.11	0.39	104
MAX3RDNOR	1.93	0.09	1.42	1.92	1.96	1.99	2	104
MEAN3RDNOR	1.02	0.16	0.61	0.92	1.04	1.11	1.46	104
MEDIAN3RDNOR	1.02	0.23	0.49	0.88	1.02	1.16	1.6	104
MIN3RDALT	0.04	0.05	0	0.01	0.03	0.06	0.24	104
MAX3RDALT	1890.44	10882.63	2.44	25.22	53.73	174.5	86545.5	104
MEAN3RDALT	37.79	208.37	0.65	2.42	3.96	7.69	1938.96	104
MEDIAN3RDALT	1.17	0.62	0.32	0.78	1.05	1.38	3.97	104
MIN3DIFFP	0.37	0.4	0	0.09	0.24	0.5	2.32	104
MAX3DIFFP	20.4	11.4	3.84	13.56	17.28	24.31	63.38	104
MEAN3DIFFP	6.93	4.04	1.55	4.74	5.94	7.75	22.8	104
MEDIAN3DIFFP	5.4	2.68	1.33	3.54	5	6.68	14.62	104

Table 18 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MIN3SPD	0.02	0.01	0	0.01	0.01	0.02	0.06	104
MAX3SPD	0.25	0.12	0.06	0.18	0.21	0.28	0.66	104
MEAN3SPD	0.13	0.05	0.03	0.1	0.12	0.15	0.3	104
MEDIAN3SPD	0.12	0.04	0.03	0.09	0.11	0.14	0.25	104
MIN3KS	9.91	4.06	3.91	7.59	9.79	11.48	30.19	104
MAX3KS	748.1	5328.42	32.93	82.68	133.82	260.68	54476.99	104
MEAN3KS	32.57	48.85	11.03	20.32	24.65	32.78	495.16	104
MEDIAN3KS	20.44	8.91	9.05	14.76	18.19	22.87	65.09	104
MIN4CV	1.54	0.97	0.12	0.85	1.34	2.04	5.53	104
MAX4CV	10.53	4.01	2.8	8.13	9.12	12.21	21.92	104
MEAN4CV	6.46	2.3	1.64	4.9	5.96	7.61	13.1	104
MEDIAN4CV	6.52	2.52	1.69	4.99	6.07	7.47	17.17	104
MIN4SKEW	-1.82	0.31	-2	-1.98	-1.94	-1.82	-0.28	104
MAX4SKEW	1.78	0.31	0.11	1.75	1.88	1.96	2	104
MEAN4SKEW	-0.06	0.5	-1.37	-0.34	-0.09	0.24	1.18	104
MEDIAN4SKEW	-0.07	0.62	-1.69	-0.38	-0.07	0.24	1.72	104
MIN4D	1.95	3.57	0.00	0.31	0.93	2.26	30.28	104
MAX4D	68.42	69.32	8.37	31.49	50.56	780.10	495.98	104
MEAN4D	25.58	33.77	2.31	10.61	17.47	25.20	274.73	104
MEDIAN4D	24.81	39.62	0.96	8.97	15.70	23.58	351.83	104
MIN4RD	0.09	0.1	0	0.02	0.05	0.11	0.45	104
MAX4RD	49.4	257.96	1.09	8.05	15.98	26.88	2627.72	104
MEAN4RD	2.33	2.67	0.35	1.15	1.76	2.51	24.7	104
MEDIAN4RD	1.24	0.92	0.16	0.74	1.07	1.35	5.95	104
MIN4RDNOR	0.13	0.14	0	0.03	0.09	0.18	0.58	104
MAX4RDNOR	2.56	0.37	1.07	2.44	2.67	2.81	3	104
MEAN4RDNOR	1.11	0.31	0.43	0.92	1.12	1.28	2.05	104
MEDIAN4RDNOR	1.07	0.4	0.22	0.84	1.08	1.25	2.27	104
MIN4RDALT	0.1	0.11	0	0.02	0.06	0.13	0.48	104
MAX4RDALT	53.65	287.9	1.1	8.68	16.19	29.49	2933.56	104
MEAN4RDALT	2.46	2.92	0.37	1.19	1.81	2.61	27.4	104
MEDIAN4RDALT	1.3	0.98	0.16	0.78	1.12	1.43	6.3	104
MIN4DIFFP	0.46	0.52	0	0.1	0.31	0.6	2.5	104
MAX4DIFFP	18.1	10.77	3.03	11.85	15.64	20.51	56.74	104
MEAN4DIFFP	6.47	4.92	1.18	3.87	5.34	6.87	26.42	104
MEDIAN4DIFFP	5.98	5.32	0.98	3.21	4.61	6.49	34.35	104
MIN4KURTO	-5.39	0.94	-6	-5.94	-5.77	-5.33	-1.12	104
MAX4KURTO	3.75	0.33	2.27	3.66	3.88	3.95	4	104
MEAN4KURTO	0.17	0.79	-1.07	-0.4	-0.06	0.56	2.41	104
MEDIAN4KURTO	0.71	0.98	-3.63	0.11	0.62	1.3	3.11	104
MIN4SPD	0.03	0.02	0	0.02	0.03	0.05	0.13	104

Table 18 continued

Predictors	Mean	Std	Min	Lower Q.	Median	Upper Q.	Max	N
MAX4SPD	0.25	0.12	0.06	0.18	0.21	0.28	0.66	104
MEAN4SPD	0.16	0.07	0.04	0.12	0.14	0.19	0.39	104
MEDIAN4SPD	0.16	0.07	0.04	0.12	0.14	0.17	0.48	104
MIN4KS	11.18	4.58	4.66	8.45	11.24	12.71	36.12	104
MAX4KS	101.52	99.91	18.25	49.53	75.09	118.27	862.88	104
MEAN4KS	22.21	10.63	9.43	16.22	19.85	25.26	97.28	104
MEDIAN4KS	18.1	7.29	5.9	13.93	17.06	20.66	59.16	104

Note: “Mean”, “Std”, “Min”, “Lower Q.”, “Median”, “Upper Q.”, “Max”, and “N” denote the mean, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and number of observations respectively. The value for “MEANBIDS”, “STDBIDS”, “D”, “MIN3D”, “MAX3D”, “MEAN3D”, ”MEDIAN3D”, “MIN4D”, “MAX4D”, “MEAN4D” and “MEDIAN4D” are expressed in thousand CHF. “KS”, “CV”, “SPD”, “RD”, “RDNOR”, “RDALT”, “SKEW”, ”DIFFP”, “KURTO”, “D”, “STDBIDS”, “MEANBIDS” and “NBRBIDS” denote the Kolmogorov-Smirnov statistic, the coefficient of variation, the spread, the relative distance, the normalized distance, the alternative relative distance, the skewness statistic, the percentage difference, the kurtosis statistic, the difference in absolute between the first and second lowest bids, the standard deviation of the bids in a tender, the mean of the bids in a tender and the number of the bids in a tender respectively. For the abbreviations, see also Table 9 in Appendix A

The normalized distance (RDNOR) assumes higher values in collusive periods than in competitive periods. For example, we find a divergence in the data between collusive tenders (with a mean of 1.38, a median of 1.24 and a standard deviation of 0.79 in Table 12) and competitive tenders (with a mean of 1.04, a median of 0.87 and a standard deviation of 0.82 in Table 13). We find similar values for the minima of the normalized distance (MIN4RDNOR) between incomplete bid-rigging cartels in sample 1 (mean of 0.37, median of 0.15, standard deviation of 0.54 in Table 14) and competitive tenders (with a mean of 0.51, a median of 0.29 and a standard deviation of 0.56 in Table 13). The values are more divergent for the maxima (MAX4RDNOR) between the two types of tender. For the incomplete bid-rigging cartels in sample 1, we observe a mean of 2.18, a median of 2.37 and a standard deviation of 0.68 in Table 14, contrasting with competitive periods, which exhibit a mean of 1.62, a median of 1.74 and a standard deviation of 0.81 in Table 13. The result indicates that the maxima of the RDNOR could be used to discriminate between incomplete bid-rigging cartels and competition.

Acknowledgements The authors would like to thank Christoph Brunner, Philippe Sulger, Niklaus Wallimann and Philipp Wegelin for support and helpful comments in the writing of this paper.

Funding Open access funding provided by University of Fribourg.

Declarations

Disclaimer All the views expressed in this paper are solely those of the authors and cannot be attributed to the Swiss Competition Commission, its Secretariat, the University of Fribourg, the University of Applied Sciences and Arts Lucerne, or Unidistance (Switzerland).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abrantes-Metz, R. M., Froeb, L. M., Geweke, J. F., & Taylor, C. T. (2006). A variance screen for collusion. *International Journal of Industrial Organization, 24*(3), 467–486.

Abrantes-Metz, R. M., Kraten, M., Metz, A. D., & Seow, G. (2012). Libor manipulation? *Journal of Banking and Finance, 36*(1), 136–150.

Aryal, G., & Gabrielli, M. F. (2013). Testing for collusion in asymmetric first-price auctions. *International Journal of Industrial Organization, 31*(1), 26–35.

Asker, J. (2010). A study of the internal organization of a bidding cartel. *American Economic Review, 100* (3), 724–762.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics, 11*, 685–725.

Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. *The Review of Economics and Statistics, 85*(4), 971–989.

Bergman, M. A., Lundberg, J., Lundberg, S., & Stake, J. Y. (2020). Interactions across firms and bid rigging. *Review of Industrial Organization, 56*(1), 107–130.

Bos, I., & Harrington, J. E. (2010). Endogenous cartel formation with heterogenous firms. *The RAND Journal of Economics, 41*(1), 92–117.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., & Cutler A. (2018). Random forests for classification and regression, cran.r-project.org

Cai, W., Mohammaditab, R., Fathi, G., Wakil, K., Ebadi, A. G., & Ghadimi, N. (2019). Optimal bidding and offering strategies of compressed air energy storage: A hybrid robust-stochastic approach. *Renewable Energy, 143*, 1–8.

Chassang, S., Kawai, K., Nakabayashi, J., & Ortner, J. (2022). Robust screens for noncompetitive bidding in procurement auctions. *Econometrica, 90*(1), 315–346.

Chernozhukov, V., Hansen, C., & Spindler, M. (2016). hdm: High-dimensional metrics. *The R Journal, 8* (2), 185–199.

Clark, R., Coviello, D., Gauthier, J.-F., & Shneyerov, A. (2018). Bid rigging and entry deterrence in public procurement: Evidence from an investigation into collusion and corruption in quebec. *The Journal of Law, Economics, and Organization, 34*(3), 301–363.

Conley, T. G., & Decarolis, F. (2016). Detecting bidders groups in collusive auctions. *American Economic Journal: Microeconomics, 8*(2), 1–38.

Crede, C. J. (2019). A structural break cartel screen for dating and detecting collusion. *Review of Industrial Organization, 54*(3), 543–574.

Decarolis, F., M. Goldmanis, and A. Penta, Marketing agencies and collusive bidding in online ad auctions, *Management Science*, 2020.

Esposito, F., and M. Ferrero, Variance screens for detecting collusion: An application to two cartel cases in italy, *Working Paper*, 2006.

Feinstein, J. S., Block, M. K., & Nold, F. C. (1985). Asymmetric information and collusive behavior in auction markets. *The American Economic Review, 75*(3), 441–460.

Foremny, A., Kulejewski, J., Anysz, H., & Nicał, A. (2018). Collusion and bid rigging in the construction industry: Case studies from poland. *Creative Construction Conference, 2018,* 822–829.

Froeb, L. M., D. S. Sibley, M. J. Doane, and B. P. Pinto, Screening for collusion as a problem of inference, *The Oxford Handbook of International Antitrust Economics*, *2*, 2014.

García Rodríguez, M. J., V. Rodríguez Montequín, F. Ortega Fernández, and J. M. Villanueva Balsera, Bidders recommender for public procurement auctions using machine learning: Data analysis, algorithm, and case study with tenders from spain, *Complexity*, *2020*, 2020.

Harrington, J. E., *Handbook of Antitrust*, 2008.

Hendricks, K., Porter, R., & Tan, G. (2008). Bidding rings and the winner's curse. *The RAND Journal of Economics, 39*(4), 1018–1041.

Ho, T. K., Random decision forests, In *Proceedings of the 3rd International Conference on Document Analysis and Recognition Montreal*, pp. 278–282, 1995.

Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization, 65,* 277–301.

Huber, M., D. Imhof, and I. Rieko, Transnational machine learning with screens for flagging bid-rigging cartels, *Journal of the Royal Statistical Society Series A*, pp. 1–41, 2022.

Hueschelrath, and Veith, Cartel detection in procurement markets, *Managerial and Decision Economics*, *35*(6), 2014.

Imhof, D., Detecting bid-rigging cartels with descriptive statistics, *Journal of Competition Law and Economics*, *15*(4), 2019.

Imhof, D., and H. Wallimann, Detecting bid-rigging coalitions in different countries and auction formats, *International Review of Law and Economics*, *68*, 106,016, 2021.

Imhof, D., Karagoek, Y., & Rutz, S. (2018). Screening for bid rigging, does it work? *Journal of Competition Law and Economics, 14*(2), 235–261.

Ishii, R. (2014). Bid roundness under collusion in Japanese procurement auctions. *Review of Industrial Organization, 44*(3), 241–254.

Jimenez, J. L., & Perdiguero, J. (2012). Does rigidity of price hide collusion? *Review of Industrial Organization, 41*(3), 223–248.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review, 105*(5), 491–95.

Liu, J., C. Chen, Z. Liu, K. Jermsittiparsert, and N. Ghadimi, An igdt-based risk-involved optimal bidding strategy for hydrogen storage-based intelligent parking lot of electric vehicles, *Journal of Energy Storage*, *27*, 2020.

McAfee, R. P., & McMillan, J. (1992). Bidding rings. *The American Economic Review, 82*(3), 579–599.

Mir, M., Shafieezadeh, M., Heidari, M. A., & Ghadimi, N. (2020). Application of hybrid forecast engine based intelligent algorithm and feature selection for wind signal prediction. *Evolving Systems, 11*(4), 559–573.

Mirzapour, F., Lakzaei, M., Varamini, G., Teimourian, M., & Ghadimi, N. (2019). A new prediction model of battery and wind-solar output in hybrid power system. *Journal of Ambient Intelligence and Humanized Computing, 10*(1), 77–87.

OECD, Roundtable on ex officio cartel investigations and the use of screens to detect cartels, 2013.

Pesendorfer, M. (2000). A study of collusion in first-price auction. *The Review of Economic Studies, 67*(3), 381–411.

Porter, R. H., & Zona, J. D. (1993). Detection of bid rigging in procurement auctions. *The Journal of Political Economy, 101*(3), 518–538.

Porter, R. H., & Zona, J. D. (1999). Ohio school milk markets: An analysis of bidding. *RAND Journal of Economics, 30,* 263–288.

Rabuzin, K., and N. Modrusan, Prediction of public procurement corruption indices using machine learning methods., *KMIS*, 2019.

Rodríguez, M. J. G., V. Rodríguez-Montequín, P. Ballesteros-Pérez, P. E. Love, and R. Signor, Collusion detection in public procurement auctions with machine learning algorithms, *Automation in Construction*, *133*, 104,047, 2022.

Silveira, D., Vasconcelos, S., Resende, M., & Cajueiro, D. O. (2022). Won't get fooled again: A supervised machine learning approach for screening gasoline cartels. *Energy Economics, 105,* 105–711.

Uslu, N. C., and F. Akal, A machine learning approach to detection of trade-based manipulations in Borsa Istanbul, *Computational Economics*, pp. 1–21, 2021.

van der Laan, M., E. C. Polley, and A. E. Hubbard, Super learner, *Statistical Applications of Genetics and Molecular Biology*, *6*(1), 2008.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance, 16*(1), 8–37.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.