Machine learning with screens for detecting bid-rigging cartels[☆]

Martin Huber [a], David Imhof [a,b,c,d,*]

[a] *Department of Economics, University of Fribourg, Switzerland*
[b] *University of Bourgogne Franche-Comté, CRESE EA 3190, Besançon F-25000, France*
[c] *Unidistance, Switzerland*
[d] *Swiss Competition Commission COMCO*

A R T I C L E   I N F O

A B S T R A C T

We combine machine learning techniques with statistical screens computed from the distribution of bids in tenders within the Swiss construction sector to predict collusion through bid-rigging cartels. We assess the out of sample performance of this approach and find it to correctly classify more than 84% of the total of bidding processes as collusive or non-collusive. We also discuss tradeoffs in reducing false positive vs. false negative predictions and find that false negative predictions increase much faster in reducing false positive predictions. Finally, we discuss policy implications of our method for competition agencies aiming at detecting bid-rigging cartels.

© 2019 Published by Elsevier B.V.

1. Introduction

In competition policy, screens are specific indices derived from the bidding distribution in tenders for distinguishing between competition and collusion as well as for flagging markets and firms likely characterized by collusion. They are thus of interest for competition agencies in order to detect cartels and to enforce competition laws. In the light of the vast variety of screens proposed in the literature (see Harrington, 2008; Jimenez and Perdiguero, 2012; OECD, 2013; Froeb et al., 2014), the question arises which detection method some competition agency should choose in practice. However, very few papers, if any, systematically investigate the performance of the screens based on statistical methods.

In this paper, we combine machine learning techniques with several screening methods for predicting collusion. We evaluate the out of sample prediction accuracy in a data set of 584 tenders that are representative for the construction sector in Switzerland. The data cover 4 different bid-rigging cartels and comprise both collusive and competitive (post-collusion) tenders, based on which we define a binary collusion indicator that serves as dependent variable. More concisely, we consider the screens proposed by Imhof (2017b) for detecting bid-rigging cartels and investigate the performance of machine learning techniques when using these screens as predictors. Firstly, we investigate lasso logit regression, (see Tibshirani, 1996), to predict collusion as a function of the screens as well as their interactions and higher order terms. Secondly, we apply an ensemble method that consists of a weighted average of predictions based on bagged regression trees (see Breiman, 1996), random forests (see Ho, 1995; Breiman, 2001), and neural networks (see McCulloch and Pitts, 1943; Ripley, 1996).

We use cross validation to determine the optimal penalization in lasso regression as well as the optimal weighting in the ensemble method. We randomly split the data into training and test samples and estimate the model parameters in the training data, while out of sample performance is assessed in the test data. We repeat these steps 100 times to estimate the average mean squared errors and classification errors. The latter is defined by the mismatch of actual collusion and predicted collusion, which is 1 if the algorithm predicts the collusion probability to be 0.5 or higher and 0 otherwise.

In our analysis, we distinguish between false positive and false negative predictions. A false positive implies that the machine learning algorithm flags a tender as collusive even though no collusion occurs. From the perspective of a competition agency, this might appear to be the worst kind of prediction error, as it could induce an unjustified investigation. In contrast, a false negative implies that the method does not flag a tender as collusive, although collusion occurs. This is undesirable, too, because any method that produces too many false negatives appears not worth being implemented due to a lack of statistical power in detecting collusion. A method that is attractive for competition agencies therefore needs to have an acceptable overall out of sample performance that satisfactorily trades off false positive and false negative prediction rates.

Our results suggest that the combination of machine learning and screening is a powerful tool for detecting bid rigging. Both the lasso and the ensemble method correctly predict 84% of all tenders out of sample or put differently, the overall misclassification rate of either method is 16%. The lasso performs slightly better in collusive than in noncollusive periods with 14% false negative and 18% false positive predictions, while the opposite holds for the ensemble method with 17% false negative and 15% false positive predictions. Therefore, the methods proposed in this paper exhibit a decent performance as the shares of false positives and negatives are low. Furthermore, we also consider reducing the share of false positives by tightening the rule for classifying tenders as collusive, i.e. by predicting collusion only to be 1 if the collusion probability predicted by the algorithm exceeds a specific threshold higher than the default value of 0.5. This exercise appears interesting for competition agencies to find an optimal tradeoff between false positives and false negatives by gauging the choice of the probability threshold. In our data, we not very surprisingly find that reducing false positives induces a considerable increase in false negative predictions.

As lasso is a variable selection method (based on constraining the sum of the absolute values of the estimated slope coefficients) for picking important predictors, it allows determining the most powerful screens. Among a set of 65 predictors that consist of the original screens as well as their interactions and squared terms, we find that two screens play a major role for detecting bid-rigging cartels: the ratio of the price difference between the second and (winning) first lowest bids to the average price difference among all bids and the coefficient of variation of bids in a tender. By far less important predictors are the number and the skewness of bids. However, we also find that discarding the two most powerful screens does not significantly affect the accuracy of the prediction, as other screens "step in" as substitutes.

As a policy recommendation, we propose a two-step procedure to detect bid-rigging cartels. The first step relies on our combination of machine learning and screening. Competition agencies may calculate the screens for each tender from the distribution of submitted bids, an information typically available in procurement processes. They may then apply the model based on screening suggested in our paper to predict collusive and competitive tenders. Concerning classification into collusive and competitive tenders, it seems (at least in our data) advisable to use a decision rule that is based on a predicted probability threshold between 0.5 to 0.7. The second step consists of scrutinizing tenders flagged as collusive by machine learning. Following Imhof et al. (2018), competition agencies should investigate if specific groups of firms or regions can be linked to the suspicious tenders. In particular, agencies can apply the bid rotation screen, see Imhof et al. (2018), which investigates the interaction among suspected firms, to check whether their group-specific interactions match a bid-rigging behavior.

Our paper is related to a small literature on implementing screens to detect bid-rigging cartels (see Feinstein et al., 1985; Imhof et al., 2018; Imhof, 2017b). This literature differs from the majority of studies on detecting bid-rigging cartels that use econometric tests typically not only relying on bidding information, but also on proxies for the costs of the

firms (see Porter and Zona, 1993; 1999; Pesendorfer, 2000; Bajari and Ye, 2003; Jakobsson, 2007; Aryal and Gabrielli, 2013; Chotibhongs and Arditi, 2012a; 2012b; Imhof, 2017a). Finally, our paper is also related to studies on screens in markets not characterized by an auction process (see Abrantes-Metz et al., 2006; Esposito and Ferrero, 2006; Hueschelrath and Veith, 2011; Jimenez and Perdiguero, 2012; Abrantes-Metz et al., 2012).

The remainder of the paper is organized as follows. Section 2 reviews our data that includes four bid-rigging cartels in the Swiss construction sector and discusses the screens used as predictors for collusion. Section 3 presents the machine learning techniques along with the empirical results. Section 4 discusses several policy recommendations of our method. Section 5 concludes.

2. Bid-rigging cartels and data

2.1. Sample description

Our data include 584 tenders with 3799 bids for a market volume of roughly 370 million Swiss francs and contain information about four different bid-rigging cartels in Switzerland. The first cartel, denoted as cartel A, was formed in the canton of Ticino (see Imhof, 2017b), the second one, denoted as cartel B, in the canton of St. Gallen (see Imhof et al., 2018). The Swiss Competition Commission (hereafter: COMCO) rendered a decision for bid-rigging cartel B but four firms appealed the decision.[1] The third and fourth cartels are denoted by C and D and their data had not been considered prior to the present paper. For confidentiality reasons, we do not report more detailed information on cartels C and D. All data on the four cases come from official records on the bidding processes at the cantonal level.

The four bid-rigging cartels concerned road construction and maintenance as well as related engineering services. More specialized engineering services like bridge or tunnel construction are, however, not included in the data. Even though the four cartels were formed in different cantons of Switzerland, the structure of the construction sector is quite comparable across cantons. Therefore, the contracts included in the data are representative for the whole of Switzerland.

For each of the cartels in our data, we observe collusive cartel periods as well as competitive post-cartel periods. Table 1 reports the number of tenders by cartel and period. Firms rigged all tenders in cartel periods and all firms submitting bids in cartel periods participated in the bid-rigging cartel. Therefore, when we subsequently refer to tenders in cartel periods, it is implied that the bid-rigging cartels are complete (also called all-inclusive) in the sense that all firms participating in the tender process were colluding. Furthermore, the firms were successful in the sense that they adhered to their

---

[1] See https://www.weko.admin.ch/weko/fr/home/actualites/communiques-de-presse/nsb-news.msg-id-64011.html.

Table 1
Number of collusive and competitive tenders.

	Collusive tenders	Perc. (%)	Competitive tenders	Perc. (%)	Total
Cartel A	148	82	33	18	181
Cartel B	19	50	19	50	38
Cartel C	93	35	174	65	267
Cartel D	39	40	59	60	98
Total	299	51	285	49	584

agreements. The opposite holds for all tenders in post-cartel periods of our data, in which firms competed to win contracts, since the selected tenders in post-cartel periods took place after the collapse of bid-rigging cartels or after the opening of an investigation.

Our data therefore offer a near-ideal setting as a laboratory of detecting bid-rigging cartels. On the one hand, the completeness of the cartels excludes the possibility of a collapse caused by the presence of non-cartel firms. If it remained unnoticed, such a collapse could entail misclassifying actually competitive tenders as collusive. On the other hand, the use of data posteriori to the collapse of the bid-rigging cartels or to the opening of an investigation should ensure that tenders are actually competitive. To sum up, we have an uncontaminated sample in the sense of having either periods of undisputed collusion or undisputed competition for evaluating the performance of simple screens in our data.

Collusive agreements were comparable across the four bid-rigging cartels and can be described as a two-step procedure. The first step consists of determining the designated winner of the tender by the cartel. Various factors play a role for how contracts are distributed among firms in a cartel, namely the distance between firms and the contract location, capacity constraints, and specialization in terms of competencies (see Imhof, 2017b, and the cartel convention of the Ticino cartel, which listed the main criteria for contract allocation). Contract allocation has to be beneficial to all in the sense that all firms should win contracts, otherwise some firms would not have an incentive to participate in bid-rigging cartels. The second step consists of determining the price of the designated winner by the cartel. This is crucial because the cover bids should be higher than the bid of the designated winner to ensure contract allocation as intended by the cartel. In other words, all firms know the price at which the designated winner of the cartel submits the bid. Even if cartel members communicate with each other on a regular base, not all members necessarily submit bids in each tender. Therefore, we frequently observe different constellations of firms for distinct tenders in our data.

In all four cartels, the procurement procedure was based on a first-price sealed bid auction. The procurement agency announced a deadline for submitting bids for a particular contract and provided all relevant documents for the tender process. Interested firms calculated and submitted their bids prior to the deadline. After the deadline passed, the call for bids was closed and the procurement agency opened the submit-

ted bids to establish a bid summary, i.e. an official record of the bid opening, which indicates the bids, the identities of the bidders, and the location and type of the contract.

Prices indicated in the bids play a major role for allocating the contracts, although procurement agencies in Switzerland take also further criteria into consideration. These include the organization of work, the quality of the solution offered by the firm, the references of the firm, and environmental as well as social aspects. Even if such additional criteria become more important as the complexity of the contracts increases, the price remains the most decisive feature in the procurement process, since the lowest bids in our data obtained the contract in most cases. Therefore, price competition was the most important dimension.[2]

Essentially two types of procedures are used by procurement agencies in Switzerland: the open procedure and the procedure by invitation. In an open procedure, all firms that meet the conditions provided in the tender documentation may submit a bid. It is legally stated that open procedures should be used for contracts above 500'000 CHF. In contrast, in the procedure by invitation, the procurement agency determines potential bidders by inviting a subset of firms (at least 3 firms, but generally more). Contracts above 500'000 CHF cannot be tendered based on invitation. Thus, competition might vary depending on the type of procedure. Since information on the latter is only partially available from bid summaries in our data, we include the number of bidders and the mean value of submitted bids in a tender (to proxy contract value) as predictors in the empirical analysis in order to account for potential differences across types of procedure.

In the paper, we only use information on bids coming from the official records of the bid opening to calculate the screens for each tender. Since access to the bid summaries is either publicly granted or easily established through procurement agencies, screening can be organized in a rather discrete manner. That is, competition agencies can conduct the screening process without attracting the attention of the bid-rigging cartel, which is crucial for any detection method.

2.2. Screens

A screen is a statistical tool to verify whether collusion likely exists in a particular market flagging unlawful behavior through economic and statistical analysis. Using a broader definition, screens comprise all methods designed to detect markets, industries, or firms for further investigation associated with an increased likelihood of collusion (see OECD, 2013). The literature typically distinguishes between behavioral and structural

---

[2] In tenders in which not only the price, but also qualitative criteria play a decisive role, another valuable information for the allocation of contracts could be the score assigned by the procurement agency to each of the bids after analysing them in detail w.r.t. the various criteria. This could be the case for complex construction projects not included in our data, like bridges, tunnels, or other specific services. However, scores are typically not publicly available in official records on bid openings, as procurement agencies analyse and rank the bids only after the opening.

screens (see Harrington, 2008; OECD, 2013). Behavioral screens aim to detect abnormal behavior of firms whereas structural screens investigate the characteristics of entire markets that may favor collusion, in order to indicate if an industry is likely prone to collusion.

Behavioral screens are divided into complex and simple methods. Complex methods generally use econometric tools or structural estimation of auction models to detect suspicious outcomes (see Porter and Zona, 1993; 1999; Baldwin et al., 1997; Pesendorfer, 2000; Bajari and Ye, 2003; Banerji and Meenakshi, 2004; Jakobsson, 2007; Chotibhongs and Arditi, 2012a, 2012b; Aryal and Gabrielli, 2013; Imhof, 2017a). Simple screens analyze strategic variables as prices and market shares to determine whether firms depart from competitive behavior. While there are many applications of simple screens to various regular markets (see Abrantes-Metz et al., 2006; Esposito and Ferrero, 2006; Harrington, 2008; Hueschelrath and Veith, 2011; Jimenez and Perdiguero, 2012; Abrantes-Metz et al., 2012; Froeb et al., 2014, for some examples), applications to bid-rigging cases are rather rare (see Feinstein et al., 1985; Imhof et al., 2018; Imhof, 2017b, for exceptions).

In this paper, we propose the application of simple screens combined with machine learning to detect bid-rigging cartels. Following Imhof (2017b), we consider several statistical screens constructed from the distribution of bids in each tender to distinguish between competition and collusion. Because each screen captures a different aspect of the distribution of bids, the combined use of different screens potentially allows accounting for different types of bid manipulation. In the following, we present the screens used as predictors in the empirical analysis.

First, bid rigging may affect the dispersion of bids in tenders due to coordination. We thus consider the standard deviation of the bids in some tender as screen. Secondly, we also include the coefficient of variation as related screen, which is defined as the ratio of the standard deviation and the average of bids:

$$CV_t = \frac{s_t}{\mu_t}, \tag{1}$$

where s_t and μ_t are the standard deviation and mean of the bids, respectively, in some tender t.

We also suspect that bid rigging affects the convergence of bids through coordination. We therefore consider the following kurtosis statistic as screen:

$$Kurt(b_t) = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{b_{it} - \mu_t}{s_t} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}, \tag{2}$$

where n denotes the number of bids in some tender t, b_{it} the ith bid, s_t the standard deviation of bids, and μ_t the mean of bids in tender t.[3]

---

[3] As the kurtosis can only be calculated for tenders with more than 3 bids, our data contain only tenders with 4 or more bids.

Furthermore, it appears plausible that bid-rigging cartels manipulate the difference between the first and second lowest bids for allocating contracts to the designated winner by the cartel. To analyze if differences between the second and first lowest bids in tenders matter, we calculate the percentage difference using the following formula:

$$Diff.Perc._t = \frac{b_{2t} - b_{1t}}{b_{1t}}, \tag{3}$$

where b_{1t} is the lowest bid and b_{2t} the second lowest bid in tender t. As an alternative screen, we also consider the absolute difference between the first and second lowest bids $b_{2t} - b_{1t}$ in the empirical analysis.

Moreover, bid manipulation may affect the symmetry of the distribution of bids in a tender. We therefore include the following skewness statistic as screen:

$$Skew(b_t) = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{b_{it} - \mu_t}{s_t} \right)^3, \tag{4}$$

where n denotes the number of the bids in some tender t, b_{it} the ith bid, s_t the standard deviation of the bids, and μ_t the mean of the bids in tender t.[4]

Bid manipulation in bid-rigging cartels might simultaneously affect the difference between the second and first lowest bids and the differences among losing bids. To capture such effects, we follow Imhof et al. (2018) and construct a relative difference ratio by dividing the difference between the second and first lowest bids by the standard deviation of all losing bids.

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

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. A relative difference higher than 1 indicates that the difference between the second and first lowest bids is greater than one standard deviation among losing bids indicating a (more) asymmetric distribution of the bids.

In addition, we normalize the difference between the first and second lowest bids in a tender dividing it by the mean of the differences between all (in terms of value) adjacent bids. We henceforth call this screen the normalized distance:

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

where n is the number of bids, b_{1t} is the lowest bid, b_{2t} the second lowest bid, and b_{it}, b_{jt} are adjacent bids in some tender t, with the values of bids being ordered increasing

---

[4] The skewness can only be computed for tenders with more than 2 bids. Our data contains only tenders with 4 or more bids.

Table 2
Descriptive statistics.

Screens	Mean	Mean	Std	Min	Max	Obs
Cartel Periods						
Coefficient of variation	3.42	2.97	1.73	0.69	13.74	299
Kurtosis	1.50	1.19	2.22	−4.03	8.14	299
Perc. difference	3.92	3.96	2.37	0.43	24.93	299
Skewness	−0.58	−0.60	1.05	−2.76	2.21	299
Relative distance	2.69	1.58	2.95	0.11	23.03	299
Normalized distance	2.23	1.96	1.31	0.19	6.95	299
Number of bidders	6.74	6.00	2.35	4	13	299
Contract value	0.731	0.381	0.868	0.023	4.967	299
Standard deviation	0.022	0.012	0.024	0.0007	0.136	299
Absolute difference	0.028	0.011	0.035	0.0003	0.272	299
Competitive periods						
Coefficient of variation	8.05	7.16	4.48	1.49	34.14	285
Kurtosis	0.07	−0.15	1.86	−5.40	6.06	285
Perc. difference	4.89	3.52	4.87	0.03	38.93	285
Skewness	0.27	0.25	0.84	−1.83	2.36	285
Relative distance	0.83	0.55	1.05	0.01	8.07	285
Normalized distance	1.01	0.85	0.74	0.01	3.77	285
Number of bidders	6.26	6.00	2.02	4	13	285
Contract Value	0.627	0.392	0.767	0.044	7.385	285
Standard deviation	0.050	0.028	0.101	0.002	1.499	285
Absolute difference	0.026	0.012	0.041	0.0001	0.323	285

Note: "Std", "Min", "Max", and "Obs" denote the standard deviation, minimum, maximum, and number of observations, respectively.

order. A value larger than 1 indicates that the difference between the second and first lowest bids is larger than the average difference between all adjacent bids in a tender.

As mentioned and discussed in the previous section, also the number of bidders and the mean of the bids submitted in a tender are included in addition to the screens, yielding all in all 10 predictors.

2.3. Descriptive statistics for the screens

Table 2 provides descriptive statistics for all screens, separately for collusive and competitive periods. We see that most of the screens generally differ in terms of means and standard deviations across both groups of periods. The mean of the coefficient of variation amounts to 3.42 in cartel periods which more than doubles in competitive periods (8.05). Moreover, the spread of the coefficient of variation is lower in cartel periods with a standard deviation of 1.73, compared to 4.48 for competitive periods. Bids are thus more similar in collusive than in competitive periods. The same applies to the standard deviation of the bids. Moreover, the mean of the kurtosis is 1.50 in cartel periods and 0.07 in competitive periods. This points to a more compressed distribution of bids in

Table 3
Statistical tests for the screens.

Screens	z-statistic	p-value MW	KSa	p-value KS
Coefficient of variation	16.30	< .0001	8.05	< .0001
Kurtosis statistic	−7.91	< .0001	3.42	< .0001
Percentage difference	0.10	0.9211	2.14	0.0002
Skewness statistic	9.69	< .0001	4.67	< .0001
Relative distance	-12.10	< .0001	5.24	< .0001
Normalized distance	−12.78	< .0001	5.63	< .0001
Number of bidders	−2.20	0.0279	1.20	0.1146
Contract value	−0.55	0.5854	0.85	0.4596
Standard deviation	8.28	< .0001	3.75	< .0001
Absolute difference	−1.30	0.1951	1.33	0.0596

Note: "Screens", "z-statistic", "p-value MW" denote the screens tested, the z-statistic of the Mann-Whitney test and the p-value of the Mann-Whitney test, respectively. "KSa" and "*p*-value KS" denote the asymptotic Kolmogorov–Smirnov statistic and the *p*-value of the Kolmogorov–Smirnov test, respectively.

collusive periods than in competitive periods, meaning that bids converge when cartel members rig contracts.

Turning to skewness, we find a mean of −0.58 in cartel periods, i.e. a tendency to left skewed distributions, contrasting with a mean of 0.25 in competitive periods. This indicates that bid rigging transforms the distribution of the bids in a more asymmetric distribution than in the competitive periods. Furthermore, the means of the relative and normalized distances amount to 2.69 and 2.23, respectively, in cartel periods, but to 0.83 and 1.01, respectively, in competitive periods. Under collusion, the difference between the first and second lowest bids tends to increase relative to differences among the total of (losing) bids. For the percentage difference between the first and the second lowest bids, the relative changes across periods are smaller: The mean and median amount to 3.92 and 3.96, respectively, for cartel periods and to 4.89 and 3.52, respectively, for competitive periods. However, the standard deviation of the percentage difference in cartel periods (2.37) is only half of that in competitive periods (4.87), pointing to a potential manipulations of the difference between the first and the second lowest bids. Concerning the absolute difference, we find no suspicious changes in the mean or in the spread across periods. Finally, the contract value exhibits a mean of 0.73 and a median of 0.38 million CHF in cartel periods, which is not too different from the respective values of 0.63 and 0.39 million CHF in competitive periods. In both types of periods, the number of bidders lies between 6 and 7 on average, but may vary between 4 to 13.

Table 3 reports Mann-Whitney and the Kolmogorov-Smirnov tests for the predictors. According to either test, the differences observed between collusive and competitive periods are statistically significant at 1% level for the coefficient of variation, the kurtosis statistic, the skewness statistic, the relative distance, the normalized distance and the standard deviation. The number of bidders is significantly different at 5% level according to the Mann-Whitney test but not to the Kolmogorov-Smirnov test. The difference in the percentage difference is significant at 1% level according to the Kolmogorov-Smirnov

test. Finally, the contract value (approximated by the mean of the bids) and the absolute difference between the first and second lowest bids are not significantly different at the 5% level. All in all, the numerous statistically significant results point to the potential usefulness of screens for predicting collusion. Note that for the lasso procedure discussed in the next section, we in additional to the 10 original predictors include interaction and second order terms for the predictors so that all in all 65 variables are used.

3. Empirical analysis using machine learning

We apply machine learning methods to train and test models for predicting bid-rigging cartels based on the screens presented in the previous section. Specifically, we consider two approaches: Lasso regression (see Tibshirani, 1996) for logit models and a so called ensemble method that consists of a weighted average of several algorithms, in our case bagged regression trees (see Breiman, 1996), random forests (see Ho, 1995; Breiman, 2001), and neural networks (see McCulloch and Pitts, 1943; Ripley, 1996).

3.1. Lasso regression

We subsequently discuss prediction based on lasso logit regression as well as the evaluation of its out of sample performance. First, we randomly split the data into two subsamples. The so-called training sample contains 75% of the total of observations and is to be used for estimating the model parameters. The so-called test sample consists of 25% of the observations and is to be used for out of sample prediction and performance evaluation. After splitting, the presence of a cartel is estimated in the training sample as a function of a range of predictors, namely the original screens as well as their squares and interaction terms to allow for a flexible functional relation.

Lasso estimation corresponds to a penalized logit regression, where the penalty term restricts the sum of absolute coefficients on the regressors. Depending on the value of the penalty term, the estimator shrinks the coefficients of less predictive variables towards or even exactly to zero and therefore allows selecting the most relevant predictors among a possibly large set of candidate variables. The estimation of the lasso logit coefficients is based on the following optimization problem:

$$\max_{\beta_0, \boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} \left[y_i \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \right) - log \left(1 + e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}} \right) \right] - \lambda \sum_{j=1}^{p} |\beta_j| \right\}. \quad (7)$$

β_0, $\boldsymbol{\beta}$ denote the intercept and slope coefficients on the predictors, respectively, x is 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 is a penalty term larger than zero. We use the predictors described in the previous section

and also include interaction and squared terms such that altogether 65 regressors are used in the lasso logit regression.

We use the lasso logit procedure in the "hdm" package for the statistical software "R" by Chernozhukov et al. (2016) and select the penalty term λ such that it minimizes the mean squared error, which we estimate by 15-fold cross-validation. This is performed by randomly splitting the training sample into 15 subsamples, also called folds. 14 folds are used to estimate the lasso coefficients under different candidate values for the penalty term. 1 fold is used as so-called validation data set for predicting cartels based on the different sets of coefficients related to the various penalties and for computing the mean squared error (MSE). The latter corresponds to the average of squared differences between the prediction and the actual presence of a cartel in the validation data. The role of folds is then swapped in the sense that each of them is used once as validation data set and 14 times for coefficient estimation, yielding 15 MSEs per penalty term. The optimal penalty term is chosen as the value that minimizes the average over the 15 respective MSE estimates.

In a next step, we run lasso logit regression in the (entire) training sample based on the (sample-size adjusted) optimal penalty term to estimate the coefficients. Finally, we use these coefficients to predict the collusion probability in the test sample. To assess the performance of out of sample prediction, we consider two measures: first, the MSE of the predicted collusion probabilities in the test sample and second, the share of correct classifications. To compute the latter measure, we create a variable which takes the value one for predicted collusion probabilities greater than or equal to 0.5 and zero otherwise, and compare it to the actual incidence of collusion in the test sample. We repeat random sample splitting into 75% training and 25% test data and all subsequent steps previously mentioned 100 times and take averages of our performance measures over the 100 repetitions.

3.2. Ensemble method

Prediction and performance evaluation for the ensemble method has in principle the same structure as for the lasso approach. The difference is that rather than lasso logit regression, any estimation step now consists a weighted average of bagged classification trees, random forests, and neural networks. The first two algorithms depend on tree methods, i.e. recursively splitting the data into subsamples in a way that minimizes the sum of squared differences of actual incidences of collusion from the collusion probabilities within the subsamples. Both methods estimate the trees in a large number of samples repeatedly drawn from the original data and obtain predictions of collusion by averaging over the tree (or splitting) structure across samples. However, one difference is that bagging considers all explanatory variables as candidates for further data splitting at each step, while random forests only use a random subset of the total of variables to prevent correlation of trees across samples. Finally, neural networks aim at fitting a system of functions that flexibly and accurately models the influence of the explanatory

Table 4
Performance of the lasso and ensemble methods.

	MSE	MSE.cart	MSE.comp	corr	corr.cart	corr.comp
Lasso	0.122	0.116	0.128	0.838	0.857	0.819
Ensemble	0.119	0.121	0.116	0.840	0.828	0.854

Note: "MSE", "MSE.cart", "MSE.comp" denote the mean squared errors in the total sample, in cartel periods, and in periods with competition, respectively. "corr", "corr.cart", and "corr.comp" denote rates of correct classification in the total sample, in cartel periods, and in periods with competition, respectively.

variables on collusion. We note that for these three machine learning algorithms, which are explained in more detail in the appendix, no higher order or interaction terms are included in addition to the original predictors, as tree-based methods are (in contrast to lasso logit) inherently nonparametric.

Cross-validation in the training sample determines the optimal weight each of the three machine learning algorithm obtains in the ensemble method, just analogously to the determination of the optimal penalty term in lasso regression. To this end we apply the "SuperLearner" package for "R" by van der Laan et al. (2007) with default values for bagged regression tree, random forest, and neural network algorithms in the "ipredbagg", "cforest", and "nnet" packages, respectively. The optimal combination of algorithms is then used to predict the collusion probabilities in the test sample and to compute the performance measures.

3.3. Empirical results

Table 4 reports the out of sample performance of lasso regression and ensemble methods in the total of the test data, as well as separately for periods with and without collusion. Both algorithms perform similarly well in terms of the MSE and the share of correctly classified cartels in out of sample data containing both cartel and post-cartel periods. Lasso exhibits a MSE of 0.12 and a correct classification rate of 84%, quasi identical to those of the ensemble method. The overall share of incorrect predictions therefore amounts to 16% for either method. When separately considering cartel and competitive periods, lasso performs slightly better for correctly classifying collusive tenders (86%) than the ensemble method (83%). The latter, however, is moderately superior for classifying competitive tenders (85%) than lasso regression (82%). Put differently, the results indicate that lasso produces 18% of false positive and 14% of false negative predictions, whereas the ensemble method produces 15% of false positive and 17% of false negative predictions.

From a policy perspective, incorrectly classifying cases of non-collusion as collusion (false positives) and thus, unnecessarily filing an investigation, might be relatively more harmful than incorrectly classifying cases of collusion as non-collusion (false negatives) and thus not detecting a subset of bid-rigging cartels. A way to reduce the incidence of incorrect classifications of actual non-collusion is raising the probability threshold for classifying a prediction as collusion from 0.5 to some higher value between 0.5 and 1.
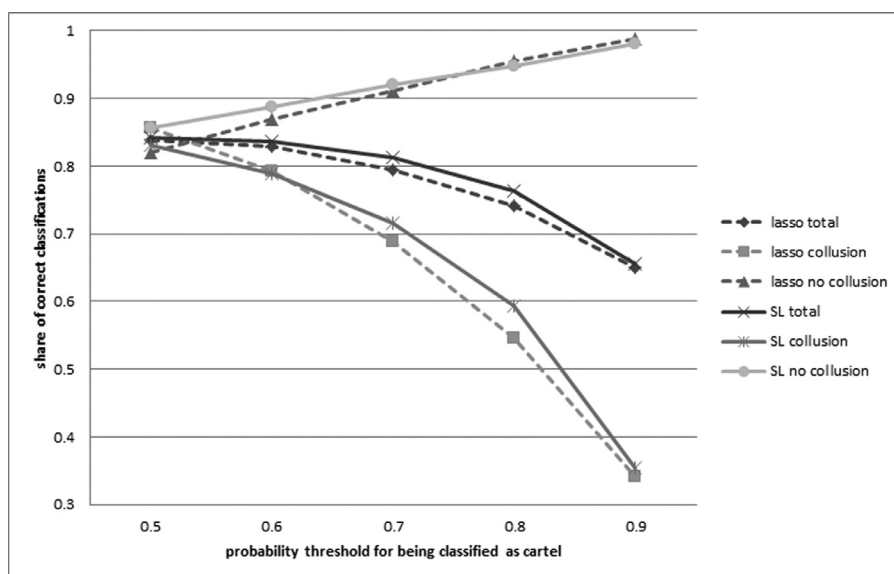
Fig. 1. False positive and false negative results by tightening the decision rule.

This, however, generally comes at the cost of reducing the likelihood of detecting actual cartels and increases therefore the false negative results. Competition agencies therefore need to appropriately trade off the likelihood of false positive and false negative results to derive an optimal rule concerning the probability threshold.

To see the tradeoffs in classification accuracy, Fig. 1 reports the correct classification rates of either methods in the total test data as well as separately for periods with and without collusion across different probability thresholds. As expected, the correct classification rate in competitive periods increases in the probability threshold for the decision rule (false positive results decrease). In contrast, the correct classification rate in cartel periods deteriorates much faster in the threshold (false negative results increase).

If a competition agency desires to minimize the risk of false positives, it could raise the probability threshold value to 0.7. In such a case, the lasso and ensemble methods classify roughly 70% of collusive tenders and more than 90% of the competitive tenders correctly, implying approximatively 10% of false positive and 30% of false negative results. The gain of reducing the risk of false positives therefore induces a disproportionate increase in false negatives. Moreover, any further tightening of the decision rule would lead to an even more severe increase of false negatives: At a probability threshold of 0.8, both methods produce 40–45% of false negatives. It therefore seems that for our data, the best-suited probability threshold lies between values of 0.5 and 0.7. One advantage of combining screening methods and machine learning consists in quantifying the trade-off of false positives and false negatives so that competition agencies are capable to determine the decision rule that optimally matches their needs.

Table 5
Average absolute values of important lasso coefficients.

Variable	NORMD	CVBID	SKEW	NBRBID
Value	1.02	0.34	0.11	0.06

Note: "NORMD", "CVBID", "SKEW", and "NBRBID" denote the the normalized difference, the coefficient of variation, the skewness, and the number of bids, respectively.

Table 6
Logit coefficients for selected screens.

	(1)	(2)	(3)	(4)
Constant	1.46***	1.56***	1.74***	−1.92***
	(0.48)	(0.41)	(0.39)	(0.20)
NORMD	1.02***	1.04***	0.89***	1.32***
	(0.21)	(0.20)	(0.15)	(0.13)
CVBID	−0.62***	−0.61***	−.61***	
	(0.06)	(0.06)	(0.06)	
SKEW	0.20	0.23		
	(0.20)	(0.18)		
NBRBID	0.02			
	(0.06)			
Obs	584	584	584	584
Pseudo-R^2	0.45	0.45	0.44	0.23

Note: "NORMD", "CVBID", "SKEW", and "NBRBID" denote the the normalized difference, the coefficient of variation, the skewness, and the number of bids, respectively. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. "Obs" and "Pseudo-R^2" denote the number of observations and the pseudo R squared.

To judge the relative importance of predictors for determining collusion, Table 5 reports the average absolute values of lasso coefficients across regressions in the 100 trainings samples (obtained by repeatedly splitting the total sample into training and test samples) that are larger than or equal to 0.05. It needs to be pointed out that in general, the estimates do not allow inferring causal associations, as lasso may importantly shrink the coefficient of a relevant predictor if it is highly correlated with another relevant predictor.[5] Nevertheless, Table 5 allows spotting the most prominently selected predictors among the total of regressors provided in the lasso regressions. We observe that the normalized difference (NORMD) and the coefficient of variation (CVBID) have by far the highest predictive power.

Table 6 reports the coefficients of standard logit regressions using various sets of screens according to their importance in lasso regression as indicated in Table 5. The coefficient of variation (CVBID) is negatively related with the probability of collusion. That is, bid rigging decreases the coefficient of variation, even conditional on the other predictors used in the four different specifications. In contrast, the normalized difference

---

[5] We also note that in our framework, causality goes from the dependent variable to the predictors rather than the other way round as it would be the case in contexts of causal inference. In causal terms, the incidence of collusion as explanatory variable affects the distribution of bids and thus the screens, which can be regarded as outcome variables. Our prediction problem therefore consists of analyzing a reverse causality: By investigating the screens, one infers the existence of their cause, namely collusion.

Table 7
Marginal effects for selected screens.

	(1)	(2)	(3)	(4)
NORMD	0.25***	0.26***	0.22***	0.33***
	(0.05)	(0.05)	(0.04)	(0.03)
CVBID	−0.15***	−0.15***	−0.15***	
	(0.01)	(0.01)	(0.01)	
SKEW	0.05	0.06		
	(0.05)	(0.05)		
NBRBID	0.01			
	(0.01)			

Note: "NORMD", "CVBID", "SKEW", and "NBRBID" denote the normalized difference, the coefficient of variation, the skewness, and the number of bids, respectively. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

(NORMD) has a (conditionally) positive association. Such results indicate that bid rigging affects the distribution of the bids in two manners in our data. First, it reduces the variance in the distribution of the bids indicating that bids are closer each other than in competitive tenders. Second, it produces asymmetry in the distribution of the bids since the differences between the second and first lowest bids are significantly higher than the differences between the losing bids in cartel periods. Finally, the coefficients of the number of bidders (NBRBIDS) are positive and therefore go against the expectation that a higher number of bidders should increase competition, but are never statistically significant at the 5% level. Skewness (SKEW) does not show any statistically significant association either, at least conditional on the other, more predictive screens.

Table 7 reports the "marginal" effects (albeit not to be interpreted causally in our predictive framework) for the average observation in the sample (i.e. at the sample means of the predictors) and various sets of screens. In accordance with the estimates in Table 6, the coefficient of variation and the normalized distance, which are significant in all models, have in absolute terms by far the largest marginal effects for explaining the probability of collusion. Augmenting the coefficient of variation by one unit decreases the probability of collusion by roughly 15 percentage points, while an increase of one unit in normalized distance raise the probability of collusion by 22 to 33 percentage points, depending on the model. The marginal effects of the other predictors are much closer to zero and not statistically significant at the 5% level.

3.4. Robustness analysis

To investigate the robustness of our methods to the omission of predictors, we discard the two most important variables, namely the normalized difference (NORMD) and the coefficient of variation (CVBID) and we apply the lasso and the ensemble method procedures to the remaining variables. Table 8 reports the MSE and the share of correctly classified cartels in out of sample data. Both the lasso and ensemble methods exhibit a correct prediction rate of 83%, implying that the remaining predictors appear to be good

Table 8
Performance when dropping 2 most important predictors.

	MSE	MSE.cart	MSE.comp	corr	corr.cart	corr.comp
Lasso	0.129	0.128	0.130	0.830	0.827	0.835
Ensemble	0.127	0.130	0.123	0.831	0.818	0.847

Note: "MSE", "MSE.cart", "MSE.comp" denote the mean squared errors in the total sample, in cartel periods, and in periods with competition, respectively. "corr", "corr.cart", and "corr.comp" denote rates of correct classification in the total sample, in cartel periods, and in periods with competition, respectively.

Table 9
Performance using only scale-invariant variables.

	MSE	MSE.cart	MSE.comp	corr	corr.cart	corr.comp
Lasso	0.122	0.115	0.129	0.840	0.862	0.818
Ensemble	0.122	0.124	0.119	0.836	0.827	0.847

Note: "MSE", "MSE.cart", "MSE.comp" denote the mean squared errors in the total sample, in cartel periods, and in periods with competition, respectively. "corr", "corr.cart", and "corr.comp" denote rates of correct classification in the total sample, in cartel periods, and in periods with competition, respectively.

substitutes for the omitted ones in our context. The results show that when important predictors are dropped, other ones become more important and "step in", such that the correct prediction rates are hardly affected. This indicates that machine learning can be quite robust to the use of a more limited set of predictors.

To investigate the robustness of our methods with respect to the choice of the screens, we apply the lasso and the ensemble procedures using only scale-invariant variables.[6] Table 9 reports the MSE and the share of correctly classified tenders in out of sample data containing both cartel and post-cartel periods. Both the lasso and the ensemble procedures exhibit correct prediction rates of nearly 84%, implying that the scale-invariant variables predict the presence or absence of bid rigging rather decently. The normalized distance (NORMD) and the coefficient of variation (CVBID) again are the most important predictors.

In the next robustness check, we change how data are split into training and test samples. Instead of randomly splitting across cartels, we train the algorithms on three cartels and use observations from the fourth cartel as test sample. In the spirit of k-fold cross-validation, each cartel serves once as test data set and three times as training data. Table 10 reports the average values for the MSE and the share of correctly classified cartels in the out of sample data. We find a correct prediction rate of 81% for the lasso and of 85% for the ensemble method. Lasso regression exhibits a correct prediction rate of 80% for collusive and of 87% for competitive periods. This contrasts with the correct prediction rates for the ensemble method, which amounts to 97% for collusive and to 75% for competitive periods. Such differences (also w.r.t. our main analysis above) are most likely explained by the unbalanced samples of collusive and competitive periods for each cartel, as shown in Table 1. Nevertheless, this robustness check by and large confirms the

---

[6] Scale-invariant variables are the coefficient of variation, the kurtosis statistic, the percentage difference, the skewness statistic, the relative difference and the normalized distance.

Table 10
Performance when using 3 cartels for training and 1 for testing.

	MSE	MSE.cart	MSE.comp	corr	corr.cart	corr.comp
Lasso	0.143	0.156	0.098	0.805	0.795	0.870
Ensemble	0.112	0.034	0.179	0.853	0.970	0.748

Note: "MSE", "MSE.cart", "MSE.comp" denote the mean squared errors in the total sample, in cartel periods, and in periods with competition, respectively. "corr", "corr.cart", and "corr.comp" denote rates of correct classification in the total sample, in cartel periods, and in periods with competition, respectively.

Table 11
Performance under misclassification of 15% of observations.

	MSE	MSE.cart	MSE.comp	corr	corr.cart	corr.comp
Lasso	0.195	0.185	0.205	0.725	0.769	0.680
Ensemble	0.189	0.182	0.197	0.735	0.744	0.726

Note: "MSE", "MSE.cart", "MSE.comp" denote the mean squared errors in the total sample, in cartel periods, and in periods with competition, respectively. "corr", "corr.cart", and "corr.comp" denote rates of correct classification in the total sample, in cartel periods, and in periods with competition, respectively.

attractiveness of machine learning, as in particular the ensemble method attains almost the same correct prediction rate as in our main analysis.

In a final check, we introduce measurement error in the outcome variable based on randomly reclassifying 15% of observations by setting collusion to competition and vice versa (using the same approach for splitting the data into training and test sets as in the main analysis). Intuitively, this introduction of measurement error is similar to an increase of the error term in a classical econometric model, implying that the signals of the predictors for classifying collusion get relatively weaker. The results are presented in Table 11. The correct prediction rates for the lasso and ensemble methods amount to 73% and 74%, respectively. This appears still satisfactory in the light of the non-negligible amount of measurement error.

4. Policy implications

Our results demonstrate the usefulness of simple screens combined with machine learning, amounting to an out of sample rate of correct classifications of 84% in our main analysis for both the lasso and the ensemble method. Furthermore, we discussed that the machine learning approach allows trading off the likelihood of false positive and false negative predictions by changing the probability threshold in a way that is considered optimal by a competition agency. This appears to be an important innovation in the literature on detecting collusion, as to the best of our knowledge no other study has directly assessed the performance of their method with respect to false positive and false negative results. We subsequently discuss some further implications of our method for policy makers, namely its advantages in terms of data requirements and use, its apparent generalizability to different empirical contexts of collusion, and its integration in a process of ex-ante detection of collusion.

4.1. Data requirements and data use

The method proposed in this paper has several advantages relative to other studies in the field. First, data requirements are comparably low. Implementing the machine learning approach and calculating the screens is straightforward and solely relies on information coming from official records of the bid opening and the bids submitted. Not even the identification of bidders is essential, which allows collecting information and implementing the method discretionarily, without attracting the attention of a cartel. This appears crucial, because if some cartel become aware of the process, it might destroy evidence such that opening an investigation would be unsuccessful. In contrast, other detection methods, such as the econometric tests proposed by Bajari and Ye (2003), require data on cost variables, which are difficult to obtain without having access to firm level data. This may compromise the secrecy in which competition agencies should implement any method aimed at detecting collusion.

We note that if the official records of the bid openings are numerous and representative for a large share of the market, it may be possible to compute cost variables for an econometric estimation of the bidding function even without directly accessing firm level data, (see Imhof, 2017a). Even in this case, the construction of the cost variables might be time-consuming (relative to the simple screens) and therefore potentially wasteful for competition agencies, which should ideally concentrate their resources on the prosecution of cartels. The resource argument is particularly relevant in the light of the high number of false negatives produced by the method of Bajari and Ye (2003) when applied to the Ticino case, see Imhof (2017a), implying that such cost variable-based approaches might have low power.

Finally, the combination of screening and machine learning allows assessing and gauging the accuracy of classifying collusion out of sample, which is relevant for data yet to be analyzed. It therefore tells us something about how well past data can be used to predict collusion in future data. To the best of our knowledge, none of the other detecting methods has so far properly assessed out of sample performance based on distinguishing between training and test data. Furthermore, by applying cross-validation to tune the algorithms, we aim at defining the best predictive model for the screens at hand, while other approaches neglect this optimization step w.r.t. model selection. Wrapping up, our paper appears to be the only one in the literature on detecting collusion that acknowledges the merits of machine learning for optimizing the predictive performance of estimators and for appropriately assessing their out of sample behavior.

4.2. Generalization of results

An important question for the attractiveness of our method is whether it yields decent results also in other contexts than that investigated in this paper. We expect our approach to perform well even in other industries or countries whenever one can ratio-

nally assume that bid rigging affects the distribution of the bids. This is even more true if the procurement procedure is similar to that considered in this paper, i.e. corresponding to a first-sealed bid auction. Furthermore, the transferability of our approach is facilitated by the use of several, distinct screens that are sensitive to different features of the distribution of bids and thus potentially cover different bid-rigging mechanisms.

Because the detection method based on simple screens is inductive, it, however, needs to be verified in the empirical context at hand. As competition agencies typically have access to data from former cases, they could easily apply the suggested screens to check their appropriateness even in different industries or countries. Moreover, if the data is large enough, an agency might directly estimate its own predictive model based on screening and machine learning to identify suspicious tenders. This is fundamentally different to establishing rigorous models for testing collusion in a deductive approach, see Bajari and Ye (2003). While deduction allows for systematic generalization if the model is correctly specified, there is the threat that a specific empirical context does not match the model parametrization and hypotheses it is based upon. At least for the Ticino case, Imhof (2017a) has shown that simple screens outperform the econometric tests proposed Bajari and Ye (2003). Combining screening with machine learning makes this flexible, data-driven approach attractive as a generally applicable tool for detecting collusion.

4.3. Ex ante detection of collusion

The trained predictive models obtained by machine learning can be applied to newly collected data in order to screen tenders for bid rigging in an ex ante procedure. Following Imhof et al. (2018), we outline possible steps of such a procedure.

Initially, a competition agency has to collect data from official records of the bid opening and compute the screens for each tender as outlined in Section 2. Applying the trained models, e.g. the lasso coefficients from the training data, to these screens allows computing the predicted collusion probabilities in the newly collected data set. Next, the competition agency needs to select a probability threshold to flag tenders as suspicious or competitive. Based on our results, we recommend to use a threshold between 0.5 and 0.7, but this could be reconsidered in other empirical contexts. We stress that the determination of suspicious tenders of such an approach is not (exclusively) based on human judgement, but data-driven and its accuracy generally improves with the amount of observations used to train the models.

If a competition agency has too few or even no data on episodes of documented cartels to train their own models for predicting bid rigging, it may instead use the model parameters estimated in this paper for detecting suspicious cases. If tenders in other jurisdictions or industries share similar structural and institutional features as our Swiss road construction data, then our trained models have the potential to work satisfactorily for detecting collusion in other countries and economic sectors, too. Whether this is in fact the case will be investigated in future research.

Once suspicious tenders have been identified, there appear two possible options concerning next steps. The first consists of immediately opening an investigation, the second one of substantiating the initial suspicion. The decision to launch an investigation should be driven by the predicted results based on machine learning. If the detection method classifies a large share of tenders in a specific period as collusive, competition agencies might want to initiate a deeper investigation immediately. For instance, a share of more than 50% may appear sufficient to inquire the opening of an investigation. In contrast, shares between 20% and 50% might seem too small to launch a (potentially costly) investigation. In this case, the market might be analyzed further to substantiate the initial suspicion.

Several approaches may be used to substantiate the initial suspicion for bid rigging. First, the firms participating in the suspicious tenders could be more closely examined to identify a specific group logic. In order to have a well-functioning bid-rigging cartel, firms must cooperate regularly over a certain period. Regular interactions between firms might make it possible to find a particular group logic in suspicious tenders. Second, geographical analysis may help identifying bid-rigging cartels situated in particular regions (see also Abrantes-Metz et al., 2006, for such a geographical screening activity). One therefore needs to determine where the suspicious tenders are localized. If they are all clustered in the same area, this might point to a local bid-rigging cartel. (Note also that if one identifies a local bid-rigging cartel, one should generally also identify a colluding group of firms based in the region.) Third, Imhof et al. (2018) assumes that bid-rigging cartels produce a rotational pattern, which can be captured with the bid rotation screen. If one is able to determine a specific group of firms regularly participating in suspicious tenders (e.g. in some region) and find that contract placement in the potential bid-rigging cartel operates in a rotational scheme, this provides further indices for substantiating the initial suspicion.

The steps proposed by Imhof et al. (2018) are not exhaustive and competition agencies might want to perform further tests and checks (e.g. according to recommendations of the OECD) to substantiate their initial suspicion. At the end of the process, the competition agencies should ideally be capable to credibly demonstrate that the suspicious bids are not coincidental, but follow an identifiable logic of collusion.

5. Conclusion

In this paper, we combined two machine learning algorithms, namely lasso regression and an ensemble method (including bagging, random forests, and neural networks), with screens for predicting collusion in tender procedures within the Swiss construction sector that are based on patterns in the distribution of bids. We assessed the out of sample performance of our approach by splitting the data into training samples for model parameter estimation and test samples for model evaluation. More than 84% of the total of bidding processes were correctly classified by both lasso regression and the ensemble methods as collusive or non-collusive.

We also investigated tradeoffs in reducing false positive vs. false negative predictions. That is, rather than classifying a tender process as collusive whenever the collusion probability predicted by machine learning is greater than or equal to 0.5, one could use a higher probability threshold for classification. This reduces incorrect predictions among truly non-collusive processes, i.e. false positives, at the cost of substantially increasing errors among truly collusive processes, i.e. false negatives. We argued that in our data, probability thresholds between 0.5 and 0.7 appear appropriate.

We found that the two most powerful screens are the coefficient of variation and the normalized distance. Conditional on other screens, the coefficient of variation is negatively and the normalized relative distance positively associated with the probability of collusion. This implies that bid rigging reduces the variance of the bids and entails an asymmetry in the distribution of bids, as the difference between the first and second lowest bids increases whereas the difference between losing bids decreases in collusive tenders of our data. However, we found that dropping these two important predictors barely affected prediction accuracy. This implies that other screens are good substitutes and "step in" to replace them. This confirms the flexibility of our approach, which does not rely on making strong assumptions about the exact relationship of one particular screen and collusion. As long as one can assume that bid rigging affects the distribution of the bids, our approach of combining machine learning and screens will likely pick this up through changes in patterns of screens.

Finally, we discussed several policy implications for competition agencies aiming at detecting bid-rigging cartels, namely advantages of our method in terms of data requirements/use, its generalizability to different empirical contexts of collusion, and its integration in a process of ex-ante detection of collusion.

Acknowledgement

Appendix A

We subsequently discuss in more detail the three machine learning approaches included in the ensemble method. Bagged trees and random forests are so-called tree-based prediction methods. The idea of a tree approach is to recursively split the predictor space in the training data, i.e. the set of possible values of the screens, into a number of nonoverlapping subregions. The incidence of collusion can then be predicted based on the share of collusive or competitive outcomes of observations that are (in terms of the values of the screens) situated in the same subregion as the observation for which the prediction should be made. The name of such methods comes from the fact that the set of splitting rules for segmenting the predictor space can be summarized in a deci-

sion tree with so-called nodes, i.e. predictor values at which the training sample is split, and leaves, which are the terminal nodes beyond which no further splitting occurs. The splitting rules could, however, also be represented in regression equations by means of dummy variables for (not) surpassing specific predictor values as well as appropriate interactions between the dummies. At each node, splitting of the predictor space is performed such that after the split, a specific goodness of fit criterion is minimized across the newly created subregions. Popular examples for such criteria are the Gini index (as it is used for binary outcomes like collusion) and the sum of squared residuals, i.e. the squared differences of the actual and average outcomes in specific subregions (used for continuous outcomes). The splitting process is repeated until a specific stopping rule is reached (e.g. maximum number of nodes or minimum number of observations in a subregion).

As with any econometric method, there exists a bias-variance trade-off in terms of model generality. Just like adding more variables or interaction terms in a regression model, more splits make the specification more flexible, which reduces bias, but entails a larger variance in the test data due to smaller subregions. A single tree with many leaves therefore likely suffers from a high variance. This issue can, however, be mitigated by bootstrap aggregation or 'bagging'. The idea is to repeatedly draw so-called bootstrap samples from the original training data with replacement (such that an observation might be drawn several times or not at all in a newly created bootstrap sample) and estimate the trees in all samples. Then, the outcome is predicted based on aggregating the predictions in the individual trees, e.g. by taking the most frequently predicted value across trees (e.g. collusion or no collusion) in case of binary outcomes, or the average of predictions in case of continuous outcomes. Random forests are similar in that they rely on creating bootstrap samples for estimating many trees and aggregating predictions. They are, however, different in that at each split, only a random subset of screens (rather than all) is chosen as potential variables for splitting. Randomly picking predictors prevents correlated trees across bootstrap samples, which generally further reduces the variance.

Neural networks are different in spirit. Rather than splitting the predictor space, they aim at fitting a system of nonlinear functions that flexibly models the influence of the predictors on collusion. Specifically, the predictors are assumed to affect specific nonlinear (conventionally sigmoid) intermediate functions, so called hidden nodes, which themselves affect the outcome of interest. The hidden nodes bear some similarity with the baseline functions in spline or series estimation, with the difference that they are learned from the data rather than predetermined, and with principal components, which are dimension-reducing linear (rather than nonlinear) functions of the predictors. Furthermore, when replacing the nonlinear functions by linear ones, neural networks collapse to a linear model, which is why they can be thought of as a nonlinear generalization of the latter. Depending on the model complexity, hidden nodes may affect the outcome either directly or through other hidden nodes, such that several layers of hidden nodes allow modelling interactions between the functions. The number of hidden nodes and

layers gauges the flexibility of the model, with more parameters reducing the bias but increasing the variance.

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, 467–486.

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

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

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

Baldwin, L.H., Marshall, R.C., Richard, J.-F., 1997. Bidder collusion at forest service timber sales. Journal of Political Economy 105, 657–699.

Banerji, A., Meenakshi, J.V., 2004. Buyer collusion and efficiency of government intervention in wheat markets in northern india: an asymmetric structural auctions analysis. American Journal of Agricultural Economics 86, 236–253.

Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.

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

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

Chotibhongs, R., Arditi, D., 2012. Analysis of collusive bidding behavior. Construction Management and Economics 30, 221–231.

Chotibhongs, R., Arditi, D., 2012. Detection of collusive behavior. Journal of Construction Engineering and Management 138, 1251–1258.

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

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

Froeb, L.M., Sibley, D.S., Doane, M.J., Pinto, B.P., 2014. Screening for collusion as a problem of inference, 2. Oxford University Press. chapter 21.

Harrington, J.E., 2008. Detecting cartel. Handbook of Antitrust. The MIT Press. Chapter 6

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

Hueschelrath, Veith, 2011. Cartel detection in procurement markets. Center for European Economic Research Discussion Paper No. 11–067.

Imhof, D., 2017a. Econometric tests to detect bid-rigging cartels: Does it work?Working Papers SES 483, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland).

Imhof, D., 2017b. Simple screens to detect bid rigging. Working Papers SES 484, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland).

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

Jakobsson, M., 2007. Bid rigging in swedish procurement auctions. Working Paper.

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

van der Laan, M., Polley, E., Hubbard, A., 2007. Super learner. Statistical Applications in Genetics and Molecular Biology 6 (1), 1–21, Article 25.

McCulloch, W., Pitts, W., 1943. A logical calculus of ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5 (4), 115–133.

OECD, 2013. Ex officio cartel investigations and the use of screens to detect cartels, Policy Roundtables. Available at: http://www.oecd.org/daf/competition/exofficio-cartel-investigation-2013.pdf

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

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

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

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58 (1), 267–288.