UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

WORKING PAPERS SES

Machine Learning with Screens for Detecting Bid-Rigging Cartels

Martin Huber
and
David Imhof

N. 494
IV.2018

Faculté des sciences économiques et sociales
Wirtschafts- und sozialwissenschaftliche Fakultät

Machine Learning with Screens for Detecting Bid-Rigging Cartels

Martin Huber* and David Imhof**

* University of Fribourg, Dept. of Economics

** Swiss Competition Commission, University of Fribourg, Dept. of Economics, University of Bourgogne Franche-Comté, CRESE, and Unidistance (Switzerland)

Abstract: *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 80% of the total of bidding processes as collusive or non-collusive. As the correct classification rate, however, differs across truly non-collusive and collusive processes, we also investigate tradeoffs in reducing false positive vs. false negative predictions. Finally, we discuss policy implications of our method for competition agencies aiming at detecting bid-rigging cartels.*

Keywords: Bid rigging detection, screening methods, variance screen, cover bidding screen, structural and behavioural screens, machine learning, lasso, ensemble methods.

JEL classification: C21, C45, C52, D22, D40, K40, L40, L41.

Addresses for correspondence: Martin Huber, University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland; martin.huber@unifr.ch. David Imhof, Hallwylstrasse 4, 3003 Bern, Switzerland; david.imhof5@gmail.com.

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 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 483 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 perform cross-validation and estimation of 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 prediction errors. 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 error rates.

Our results suggest that the combination of machine learning and screening is a powerful tool to detect bid-rigging. Lasso logit regression correctly predicts out of sample 82% of all tenders. However, the rate differs across cartel and non-cartel cases. While lasso correctly classifies 91% of the collusive tenders (i.e. 9% are false negatives), it correctly classifies 69% of the competitive tenders (31% false positives classified as collusive in the absence of bid-rigging). Thus, false positives rates are more than three times higher than false negatives. To reduce the share of false positives (which generally comes with an increase of false negatives), we consider tightening the classification rule, by only classifying a bid as collusive if the predicted collusion probability is larger than or equal to 0.7 (rather than 0.5). In this case, lasso correctly classifies 77% of collusive tenders (23% false negatives) and 85% of competitive tenders (15% false positives). By gauging the choice of the probability threshold, a competition agency may find an optimal tradeoff between false positives and false negatives.

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. We find that two screens play a major role for detecting bid-rigging cartels, namely the ratio of the price difference between the second and (winning) first lowest bids to the average price difference among all losing bids and the coefficient of variation of bids in a tender. By far less important predictors are the number and skewness of bids.

Concerning the ensemble method, we note that it very slightly dominates lasso in terms of overall performance with 83% of classifications being correct when the probability threshold is 0.5. While the correct prediction rate of the ensemble method is slightly below that of lasso for collusive tenders (88%), it is higher for competitive tenders (76%). When setting the probability threshold to 0.7, the ensemble method does slightly better than lasso both among correctly classified collusive tenders (80%) and competitive tenders (86%), but the performance of either method appears satisfactory.

As 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 tighter decision rule, by raising the probability threshold from 0.5 to 0.7. This importantly reduces the risk of false positives, at the cost of somewhat increasing the rate of false negatives. The second step consists of scrutinizing tenders flagged as collusive by machine learning. Following *Imhof et al.* (2017), competition agencies should investigate if specific groups of firms or regions can be linked to the suspicious tenders. In particular, agencies can apply the cover-bidding screen, see *Imhof et al.* (2017), 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.*, 2017; *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,b; *Imhof*, 2017a). However, such cost information is not easily available before the opening of an investigation and the data collection process in order to implement such tests (see *Bajari and Ye*, 2003) is rather complex when compared to our method based on machine learning and screening. 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, which includes four bid-rigging cartels in the Swiss construction sector. Section 3 discusses the screens used as predictors for collusion. Section 4 presents the machine learning techniques along with the empirical results. Section 5 discusses several policy recommendations of our method. Section 6 concludes.

2 Bid-Rigging Cartels and Data

In this section, we discuss our data which 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.*, 2017). The Swiss Competition Commission (hereafter: COMCO) rendered a decision for bid-rigging cartel B but four firms appealed against 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 any related further engineering services. More special engineering services as 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, in which the cartels were active, is quite comparable. Therefore, the contracts included in the data are representative for the whole of Switzerland.

For each of the cartels in our data, we observe the cartel period as well as a competitive post-cartel period. Table 1 reports the number of tenders by cartel and period. Firms rigged all tenders in the cartel period and all firms submitting bids in the cartel period participated in the bid-rigging cartel. Therefore, when we subsequently refer to tenders in the cartel period, it is implied that the

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

bid-rigging cartels are complete 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 agreements. The opposite holds for all tenders in the post-cartel periods of our data, in which firms fully competed to win contracts. Thus, we have an uncontaminated sample in the sense of having either periods of perfect collusion or perfect competition for evaluating the performance of sample screens.

Table 1: Number of collusive and competitive tenders

Cartel Period	Perc.	Post-cartel	Period	Perc.	Total
Cartel A	148	82%	33	18%	181
Cartel B	19	50%	19	50%	38
Cartel C	94	53%	85	47%	179
Cartel D	39	46%	46	54%	85
Total	300	62%	183	38%	483

Collusive agreements were comparable across the four bid-rigging cartels and can be described as a two-steps 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. Contract allocation has to be beneficial to all in the sense that all firms should win contracts, otherwise certain firms would not have incentives 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.

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 submitted 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.

In the paper, we use solely 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.

Essentially two types of procedures are used by procurement agencies at the 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 can vary depending on the type of procedure and one would suspect it to be fiercer in the open procedure than by invitation. To take account of such differences in the pressure to compete, we include the number of bidders and the value of the contract as potential predictors in the empirical analysis.

Prices indicated in the bids play a major role for allocating the contracts, although procurement agencies in Switzerland take also further criteria into consideration. This includes 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.

3 Screens

A screen is a statistical tool to verify whether collusion likely exists in a particular market and its purpose is to flag 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 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; *Baldwin et al.*, 1997; *Porter and Zona*, 1999; *Pesendorfer*, 2000; *Bajari and Ye*, 2003; *Banerji and Meenakshi*, 2004; *Jakobsson*, 2007; *Aryal and Gabrielli*, 2013; *Chotibhongs and Arditi*, 2012a,b; *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, applications to bid-rigging cases are rather rare (see *Feinstein et al.*, 1985; *Imhof et al.*, 2017; *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 in more detail present the screens used in the empirical analysis. First, we discuss two kinds of behavioral screens: variance screens and the cover-bidding screens. We consider the coefficient of variation and the kurtosis as variance screens, and the percentage difference between the first and second lowest bids, the skewness, and two measures of the difference of the second and first lowest bids relative to the differences among all losing bids as cover bidding screens. For each behavioral screen, we describe by means of the Ticino case (cartel A) how collusion affects the distribution of bids and the screens. Second, we consider two structural screens: The number of bidders per tender and the size of the contract. We discuss their possible influence on the pressure to compete, which determines the likelihood of collusion.

3.1 Variance Screens

3.1.1 Coefficient of Variation

Cartel members must exchange information in order to coordinate bids and assure that the designated winner by the cartel actually acquires the contract. That exchange of information on bids likely affects the support of the distribution of bids. First, cartel members do not make too low bids, since their aim is to raise the bid of the designated winner in order to extract a positive cartel rent. Second, cartel members cannot submit too high bids either, because procurement agencies might have a certain prior about a realistic distribution of bids. Therefore, cartel members are constrained to submit higher bids than under competition that are, however, still below a certain threshold. This reduces the support of the distribution of bids and thus, in general the variance.

To capture the effect of the support reduction in the distribution of bids, we consider the coefficient of variation represented by the following formula:

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

where s_t and μ_t are the standard deviation and mean of the bids, respectively, in some tender t. μ_t is higher in cartels than in competitive markets, since cartel members submit higher bids. Furthermore, s_t decreases because the support is reduced. Since μ_t increases and s_t decreases, CV_t necessarily decreases when bid rigging occurs. The subsequent graph illustrates this for the Ticino case, (see *Imhof*, 2017b). The vertical lines delimit the cartel period going from January 1999 to April 2005.

It can be seen that the coefficient of variation is significantly lower in the cartel period compared to the post-cartel period or the year 1998. In the cartel period, the coefficient of variation exhibits a mean and a median amounting to 3.43 and 3.13, respectively. In contrast, in the post-cartel period the mean and median are 8.92 and 8.10, respectively, which approximatively corresponds to the values prior to the cartel period (see *Imhof*, 2017b). Therefore, the coefficient of variation is expectedly lower when our dependent variable *cartel period* takes the value 1. In the empirical analysis further below, we investigate if the lasso coefficient on the predictor *coefficient of variation* is negative and

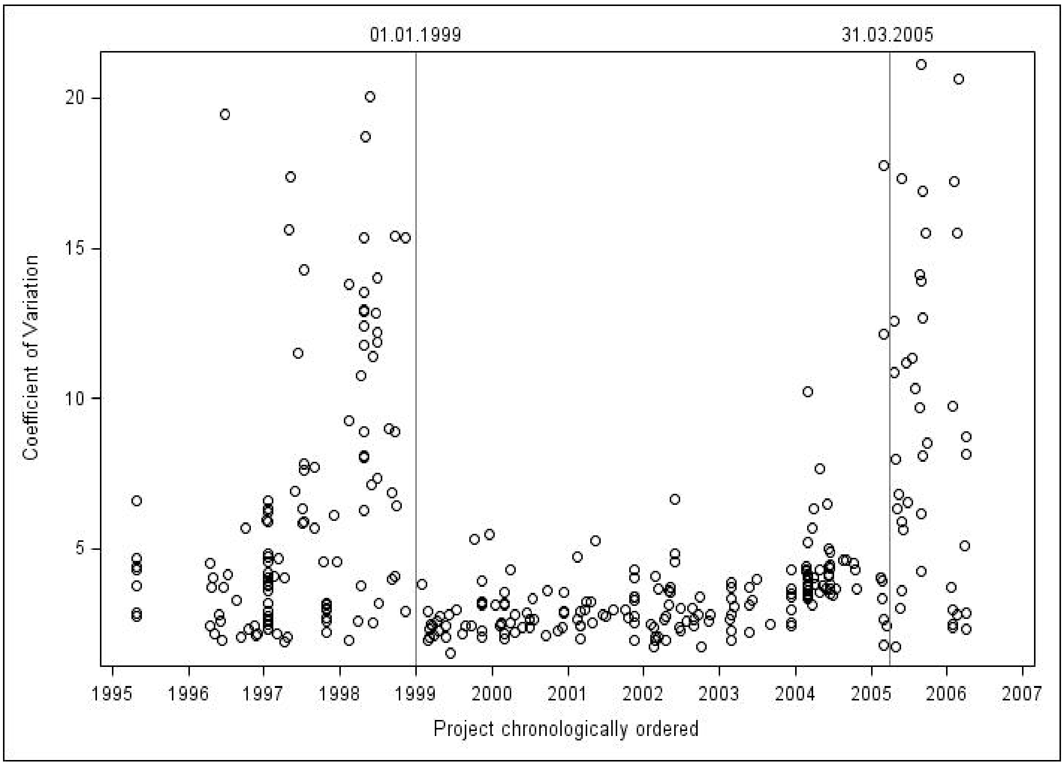

Figure 1: The Evolution of the Coefficient of Variation

how it impacts the likelihood of collusion.

Observation 1: *the coefficient of variation decreases in the case of bid rigging.*

3.1.2 Kurtosis

We suspect that bids converge when bid rigging occurs. Bidders exchange their bids, in particular that of the designated winner of the cartel. All other firms submit phony bids that use the bid of the designated winner as focal point. Because other cartel members submit phony bids slightly higher than that of the designated winner, the distribution of the bids is more compressed than in competitive markets, i.e. there is a tendency of convergence. Similar to support reduction, convergence tends to decrease the variance. Yet the two phenomena are conceptually not exactly the same: Convergence of bids implies that the mean absolute difference between bids is reduced, possibly beyond the reduction of the support of the distribution of bids. To analyze the convergence effect of bid rigging on the distribution of bids, we consider the following kurtosis statistic:

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

where n denotes the number of bids in some tender t, b_{it} the i^{th} bid, s_t the standard deviation of bids, and μ_t the mean of bids in tender t.[2] In the Ticino case, we observe that the kurtosis is positive and significantly higher in the cartel period than in the post-cartel period. During the cartel period, bids show a tendency of convergence with the mean and median kurtosis amounting to 2.71 and 2.84, respectively. In contrast, in the post-cartel period the mean and median are -0.08 and -0.16, respectively, which approximatively corresponds to the values of a normal distribution, (see *Imhof*, 2017b). Therefore, the kurtosis is expectedly positive and higher when the dependent variable *cartel period* takes the value 1.

Observation 2: *the kurtosis statistic is positive and increases in the case of bid rigging.*

3.2 Cover-Bidding Screens

3.2.1 Percentage Difference

We suspect that the difference between the first and second lowest bids matters to ensure that the designated winner by the cartel indeed wins the contract. If the second lowest bid is too close to the lowest one, the procurement agency might place the contract with the second lowest bid when criteria other than the price appear more favorable (in a way that they compensate for the price difference). As mentioned before, such criteria comprise the technical solution offered, the quality, the references of the firms, and environmental or social aspects. While for standard construction works it is the prices that are most decisive, references and quality of the technical solution may matter for more specialized works. Since our data mainly consist of standard construction works and related engineering services, price is the major criterion for awarding the contract. But even then, firms might want to maintain a certain minimum difference between the first and the second

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

lowest bids to guarantee the outcome desired by the cartel. This is decisive for the stability of the bid-rigging cartel.

To analyze the difference between the second and first lowest bids in a tender, 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.

In the Ticino case, the percentage difference lies between 2.5% and 5% during the cartel period, while it decreases in the post-cartel period and is below 2.5% for half of the observations, (see *Imhof*, 2017b). In other cartels, it has been observed that the second lowest bid is generally between 3% and 5% higher than the lowest one when collusion occurs. Therefore, we expect the percentage difference between the second and first best bids to increase when the dependent variable *cartel period* takes the value 1.

Observation 3: *The percentage difference between the two lowest bids in a tender increases in the case of bid rigging.*

3.2.2 Skewness Statistic

We suspect the distribution of bids to become (more) asymmetric in the case of bid-rigging compared to competition, due to a greater difference between the first and the second lowest bids, see the discussion of the previous section. We analyze this using the skewness statistic:

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

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

In the Ticino case, the skewness statistic is negative in the cartel period with a mean of -1.06 and

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

a median of -1.29 such that the distribution of bids is left-skewed. In contrast, the distribution is almost symmetric in the post-cartel period with the mean and median skewness amounting to 0.24 and 0.37, respectively, (see *Imhof*, 2017b). Therefore, when the dependent variable *cartel period* takes the value 1, the skewness is expectedly negative, indicating an asymmetric distribution of bids.

Observation 4: *The skewness decreases in the case of bid rigging and is negative.*

3.2.3 Relative Difference

Concerning our relative difference measure, we combine the arguments of the percentage difference and skewness. We assume simultaneously that the difference between the second and first lowest bids increases and the difference among losing bids decreases. To analyze this cover bidding behavior, we construct a relative difference ratio following *Imhof et al.* (2017) by dividing the difference between the second and first lowest bids $\Delta_{1t} = b_{2t} - b_{1t}$ by the standard deviation of all losing bids.

$$RD_t = \frac{\Delta_{1t}}{s_{t,losingbids}}, \tag{5}$$

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. This points to a cover-bidding mechanism.

For the Ticino case, we observe that the relative difference is higher than 1 in the cartel period with a mean of 4.15 and a median of 3.08, such that the differences between the second and first lowest bids are generally clearly larger than the respective standard deviations among losing bids. As shown in graph 2 below, very few values are below 1 in the cartel period, while in the post-cartel period, the mean and the median of the relative differences only amount to 0.84 and 0.62, respectively, (see *Imhof*, 2017b). Therefore, when the dependent variable *cartel period* takes the value 1, the relative distance is expectedly larger (than 1), pointing to an asymmetric distribution of bids.

Observation 5: *The relative difference increases in the case of bid rigging and is higher than 1.*

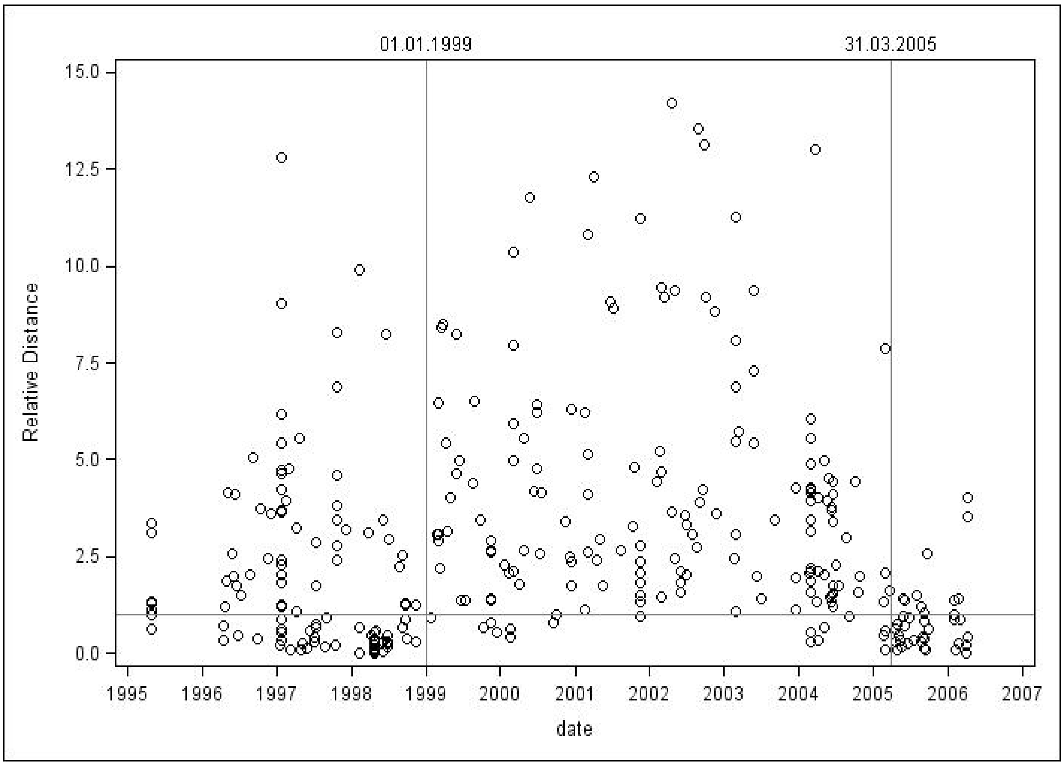

Figure 2: Evolution of the Relative Distance

3.2.4 Alternative Relative Difference

We also calculate the relative difference in an alternative way, based on the same expectations as for the previously considered relative difference: The difference between the second and first lowest bids presumably increases, while the difference among losing bids generally decreases. However, we now use the mean difference between losing bids (rather than the standard deviation) as denominator in our alternative ratio for the relative difference:

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

where n is the number of bids, b_{1t} is the lowest bid, b_{2t} the second lowest bid, and the term in bracket is the sum of differences between losing bids in some tender t. A value larger than 1 indicates that the difference between the second and first lowest bids is larger than the average difference between losing bids.

Observation 6: *The alternative relative difference increases in the case of bid rigging and is greater* *than 1.*

3.3 Structural Screens

Concerning structural screens, we consider two variables: the number of bidders per tender and the value of the contract. Both variables may influence the intensity of competition in the tender process.

3.3.1 Number of Bidders

The number of bidders supposedly affects the intensity of competition. The larger the number of bidders is, the fiercer competition is generally expected to be. The pressure to compete may also have an impact on the magnitudes of behavioral screens or interact with them. It thus appears important to control for the number of bidders in our empirical analysis.

3.3.2 Contract Value

The value of the contract presumably affects competition in a positive way. As the value of the contract increases, so does the revenue of the firm winning the contract. Therefore, we suspect firms to compete more fiercely for a contract worth, say, 2 million CHF than for 0.2 million CHF. If the value of the contract affects intensity of competition, it may interact with behavioral screens, making it a potentially important control variable in the empirical analysis.

3.4 Descriptive Statistics

Table 2 provides descriptive statistics for all screens, separately for collusive and competitive periods. We see that the screens generally differ in terms of means and standard deviations across both groups of periods. According to the Mann-Whitney test, these differences are statistically significant at the 1% level for all but the percentage difference screen. Furthermore, the Kolmogorov-Smirnov test for equality in distributions is significant at 1% for all screens. This points to the potential usefulness of the screens for predicting collusion, which is empirically assessed in the next section.

Table 2: Descriptive Statistics

Screens	Mean	Median	Std	Min	Max	Obs
Cartel Periods						
Number of bidders	6.75	6.00	2.34	4	13	300
Relative Distance	2.69	1.58	2.95	0.11	23.02	300
Alternative Relative Distance	2.23	1.98	1.31	0.19	6.95	300
Coefficient of Variation	3.53	2.99	1.96	0.69	13.74	300
Kurtosis	1.48	1.16	2.23	-4.03	8.14	300
Skewness	-0.60	-0.61	1.04	-2.76	2.2	300
Perc. Difference	4.03	4.02	2.62	0.43	24.93	300
Contract Value	730’603.68	382’491.26	866’905.88	23’159.81	4’967’503.78	300
Competitive Periods						
Number of bidders	6.16	6.00	1.98	4	13	183
Relative Distance	0.83	0.51	0.96	0.01	5.85	183
Alternative Relative Distance	1	0.83	0.73	0.01	3.67	183
Coefficient of Variation	7.94	7.23	3.94	1.49	22.59	183
Kurtosis	0.22	0.07	1.81	-5.40	6.06	183
Skewness	0.31	0.34	0.86	-1.83	2.36	183
Perc. Difference	4.90	3.30	5.18	0.03	38.93	183
Contract Value	690’019.83	458’394.79	714’775.21	43’973.44	5’180’863.95	183

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

4 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 Section 3. 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).

4.1 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:

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

β_0, β 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 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.

4.2 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 of 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 variables on collusion. We note that for these three machine learning algorithms, no higher order or interaction terms are included in addition to the original screens, as they 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.* (2008) 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.

4.3 Empirical Results

Table 3 reports the out of sample performance of the lasso 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. The ensemble method slightly dominates with an MSE of 0.12 and a correct classification rate of 83%, compared to 0.13 and 82% for lasso. When considering cartel periods only, both methods have a very similar MSE of roughly 0.08, but the lasso performs slightly better in terms of the correct classification (of collusion) with a rate of 91%, compared to 88% for the ensemble method. The latter, however, works better in non-cartel periods: the MSE and the correct classification rate (of no collusion) amount to 0.18 and 76%, respectively, while the lasso attains 0.20 and 69%. It is worth noting that both methods perform considerably better in cartel rather than competitive periods.

Table 3: Performance of The Lasso and Ensemble Methods

	MSE	MSE.cart	MSE.comp	corr	corr.cart	corr.comp
lasso	0.127	0.084	0.196	0.824	0.907	0.690
ensemble	0.120	0.082	0.181	0.834	0.881	0.760

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.

From a policy perspective, incorrectly classifying cases of non-collusion as collusion (false positives) and thus, unnecessarily filing an investigation, might be relatively more sensitive 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. 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, Figure 3 reports the correct classification rates of either method 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. The improvement is steeper for the lasso estimator, which starts out at an inferior level for a threshold of 0.5, but slightly outperforms the ensemble method after 0.8. In contrast, the correct classification rate in cartel periods deteriorates much faster in the threshold when using the lasso estimator rather than the ensemble method. For a threshold of 0.9, the rate is at 38% for the lasso (i.e. substantially worse than flipping a coin), while it is still at 59% for the ensemble method. For this reason, the overall correct classification rate in the total of the test data decreases slower for the ensemble estimator than for the lasso estimator. For a probability threshold of 0.9, the share of correct predictions amounts to 72% for the former, but just 60% for the latter method.

From a policy perspective, a probability threshold value of 0.7 for the decision rule appears to be a pertinent choice. Lasso correctly classifies 77% of the collusive tenders and 85% of the competitive tenders, whereas the ensemble method correctly classifies correctly 80% of the collusive tenders and 86% of the competitive tenders. Even if the ensemble method does slightly better than lasso, their performances are quite similar. When further tightening the decision rule by raising the probability threshold to 0.8, both methods correctly classify roughly nine out of ten competitive tenders, while false negative results increase: the ensemble algorithm still detects 3 out of 4 bid-rigging cartels when collusion is present and therefore outperforms the lasso estimator with identifying roughly 2 out 3 bid-rigging cartels. The advantage of combining screening methods and machine learning consists in quantifying the trade-off of false positives and false negatives such that competition agencies are capable to determine the decision rule that best suits their needs.

To judge the relative importance of predictors for determining collusion, Table 4 reports the

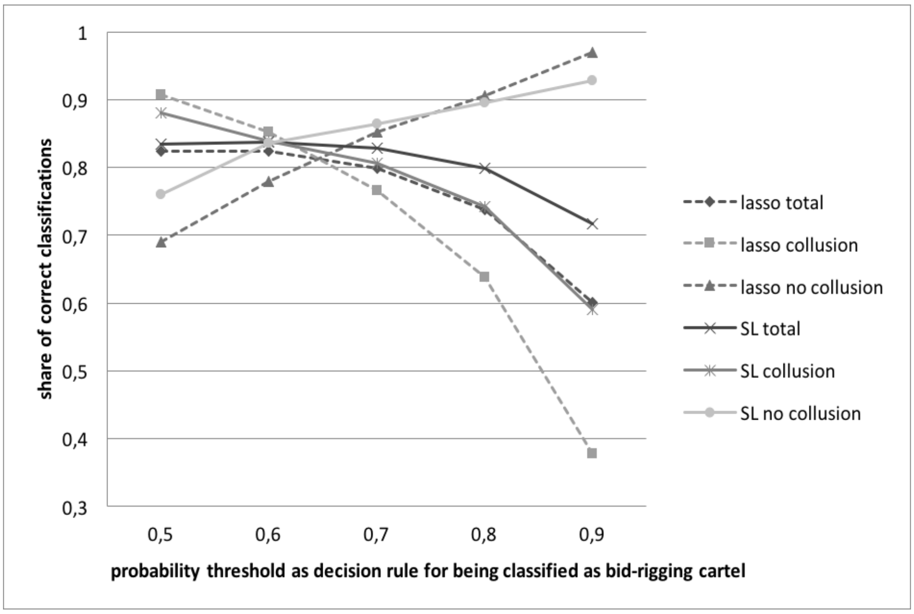

Figure 3: False Positive and False Negative Results by Tightening the Decision Rule

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 0.03. 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.[4] Nevertheless, Table 4 allows spotting the most prominently selected predictors among the total of regressors provided in the lasso regressions. We observe that the alternative relative difference (ALTRD) and the coefficient of variation (CVBID) have by far the highest predictive power.

Table 4: Average Absolute Values of Important Lasso Coefficients

ALTRD	0.699
CVBID	0.394
NBRBID	0.039
SKEW	0.035
RD	0.030

Note: “ALTRD”, “CVBID”, “SKEW”, and “RD” denote the the alternative relative difference, the coefficient of variation, the number of bids, the skewness, and the relative difference, respectively.

[4] 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, it is the incidence of collusion which 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 analysing a reverse causality: By investigating the screens, one infers the existence of their cause, namely collusion.

Table 5 reports the coefficients of standard logit regressions using various sets of screens according to their importance in lasso regression as indicated in Table 4. For the statistically significant variables, we confirm the expectations concerning the behavior of the screens inferred from the Ticino case. 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 alternative relative difference (ALTRD) has a (conditionally) positive association. This implies that the distribution of the bids converges and that the differences between the second and first lowest bids are significantly higher than the differences between the losing bids. The same pattern can be observed for the relative difference (RD), for which the association is, however, statistically insignificant. 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 5: Logit Coefficients for Selected Screens

	(1)	(2)	(3)	(4)
Constant	0.82	1.04**	1.02**	1.51***
	(0.56)	(0.51)	(0.5)	(0.4)
CVBID	-0.48***	-0.49***	-0.49***	-0.47***
	(0.06)	(0.06)	(0.06)	(0.06)
ALTRD	0.73**	0.89***	0.92***	0.95***
	(0.29)	(0.23)	(0.17)	(0.16)
NBRBIDS	0.13*	0.10	0.09	
	(0.07)	(0.06)	(0.06)	
SKEW	0.04	-0.05		
	(0.23)	(0.21)		
RD	0.17			
	(0.19)			
Obs	483	483	483	483
R^2	0.41	0.41	0.41	0.41

Note: "ALTRD", "CVBID", "SKEW", and "RD" denote the the alternative relative difference, the coefficient of variation, the number of bids, the skewness, and the relative difference, respectively. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. "Obs" and "R^2" denote the number of observations and the pseudo R squared.

Table 6 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 5, the coefficient of variation and the alternative relative 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 10 percentage points, while a one unit increase in alternative distance makes the probability increase by 16 to 20 percentage points, depending on the model. The marginal effects of the other predictor a much closer to zero and not statistically significant at the 5% level.

Table 6: Marginal Effects for Selected Screens

	(1)	(2)	(3)	(4)
CVBID	-0.10***	-0.11***	-0.11***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.01)
ALTRD	0.16**	0.20***	0.20***	0.20***
	(0.06)	(0.05)	(0.03)	(0.03)
NBRBIDS	0.03*	0.02	0.02	
	(0.02)	(0.01)	(0.01)	
SKEW	0.01	-0.01		
	(0.05)	(0.05)		
RD	0.04			
	(0.04)			

Note: “ALTRD”, “CVBID”, “SKEW”, and “RD” denote the the alternative relative difference, the coefficient of variation, the number of bids, the skewness, and the relative difference, respectively. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively.

5 Policy Implications

The results of the previous section demonstrate the usefulness of simple screens combined with machine learning, amounting to an out of sample rate of correct classifications of 82% and 83% for the considered lasso and ensemble approaches, respectively. They confirm the observations and assumptions drawn from the Ticino case for all the screens used in the regressions. 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.

5.1 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 gets aware of the process, it might destroy evidence such that opening an investigation would be unsuccessful. In contrast, other detection methods as the econometric tests proposed by *Bajari and Ye* (2003) require data on the 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 complex and 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 analysed. 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.

5.2 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 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 base 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, *Im-*

hof (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.

5.3 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.* (2017), 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 3. 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, using a threshold of 0.7 appears sensible, 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.

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. 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.* (2017) assumes that bid-rigging cartels produce a rotational pattern, which one can detect by a collusive interaction test. 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.* (2017) 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.

6 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 80% of the total of bidding processes were correctly classified by both lasso regression and the ensemble methods as collusive or non-collusive. As the correct classification rate, however, differs across truly non-collusive and collusive processes, 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, e.g. 0.7. This reduces incorrect predictions among truly non-collusive processes, i.e. false positives, at the cost of somewhat increasing errors among truly collusive processes, i.e. false negatives.

We demonstrated that setting the probability threshold to 0.7 in our data entailed acceptably low shares of both false positives and false negatives. Lasso regression correctly classified 77% of collusive tenders and 85% of competitive tenders, whereas the ensemble method correctly classified 80% of collusive tenders and 86% of competitive tenders. 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.

References

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

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

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

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

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

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

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

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

Chernozhukov, V., C. Hansen, and M. Spindler, hdm: High-dimensional metrics, *preprint arXiv:1608.00354*, 2016.

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

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

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

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

Froeb, L. M., D. S. Sibley, M. J. Doane, and B. P. Pinto, *The Oxford Handbook of International Antitrust Economics*, vol. 2, chap. 21 Screening for Collusion as a Problem of Inference, Oxford University Press, 2014.

Harrington, J. E., *Handbook of Antitrust*, chap. 6 Detecting Cartel, The MIT Press, 2008.

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

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

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

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

Imhof, D., Y. Karagoek, and S. Rutz, Screening for bid rigging, does it work?, *Working Paper UFC*, 2017.

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

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

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

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

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

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

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

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

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

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

Working Paper 494 april 2018

Authors
Martin HUBER
University of Fribourg, Department of Economics, Chair of Applied Econometrics - Evaluation of Public Policies, Bd. de Pérolles 90, 1700 Fribourg, Switzerland.
Phone: +41 26 300 8274; Email: martin.huber@unifr.ch; Website: http://www.unifr.ch/appecon/en/team/martin-huber

David IMHOF
Economist, Secretariat of the Swiss Competition Commission. Email: david.imhof@weko.admin.ch
Affiliated to the University of Fribourg (Switzerland), the University of Franche-Comté (France) and Unidistance (Formation universitaire à distance, Switzerland)

Abstract
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 80% of the total of bidding processes as collusive or non-collusive. As the correct classification rate, however, differs across truly non-collusive and collusive processes, we also investigate tradeoffs in reducing false positive vs. false negative predictions. Finally, we discuss policy implications of our method for competition agencies aiming at detecting bid-rigging cartels.

Citation proposal
Martin Huber, David Imhof. 2018. «Machine Learning with Screens for Detecting Bid-Rigging Cartels». Working Papers SES 494, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland)

Jel Classification
C21, C45, C52, D22, D40, K40, L40, L41.

Keywords
Bid rigging detection, screening methods, variance screen, cover bidding screen, structural and behavioural screens, machine learning, lasso, ensemble methods.

Publisher
Université de Fribourg, Suisse, Faculté des sciences économiques et sociales
Universität Freiburg, Schweiz, Wirtschafts- und sozialwissenschaftliche Fakultät
University of Fribourg, Switzerland, Faculty of Economics and Social Sciences

Bd de Pérolles 90, CH-1700 Fribourg
Tél.: +41 (0) 26 300 82 00
decanat-ses@unifr.ch www.unifr.ch/ses