DETECTING SUSPICIOUS BEHAVIOR IN PUBLIC PROCUREMENT AUCTIONS

by

Maryna Tonkoshkur

A thesis submitted in partial fulfillment of the requirements for the degree of

MA in Economic Analysis

.

Kyiv School of Economics

2018

Thesis Supervisor: ______________________________ Professor Elena Besedina

Approved by ____
Head of the KSE Defense Committee, Professor Tymofiy Mylovanov

____

____

____

Date ____

Kyiv School of Economics

Abstract

DETECTING SUSPICIOUS
BEHAVIOR IN PUBLIC
PROCUREMENT AUCTIONS

by Maryna Tonkoshkur

Thesis Supervisor: Professor Elena Besedina

This purpose of this work is to build a model to assess the factors, which are associated with the suspicious behavior of firms in the Ukrainian public procurement auctions. The dependent variable in this analysis is constructed using the sequential pattern mining algorithm and indicator of the constant bids during all rounds of auction. Next, the logistic regression is used to evaluate lots and participants characteristics associated with suspicious behavior. The main data source is the electronic public procurement system "ProZorro" with information on tenders held in 2016 for three categories of goods: Agricultural and related products; Petroleum and other sources of energy; Food and related products.

The results of estimations suggest that firms with suspicious behavior in agricultural and food industries tend to choose the lots with the estimated value, which is slightly above the threshold of 200 thousand UAH or significantly below the threshold value in the case of open competitive procedures. Also in those industries suspicious firms almost never participate in lots only with members of bidding ring. However, in the industry of "Petroleum products, fuel, electricity and other sources of energy" suspicious firms tend to choose below threshold lots with higher estimated values. The entities, which were previously investigated by Antimonopoly Committee of Ukraine tend to possess more suspicious behavior only in Petroleum

products, fuel, electricity and other sources of energy industry, while in the other two industries such firms are less suspicious.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The author would like to express her sincere gratitude to her thesis advisor Professor Elena Besedina for the inspiration to dig into the topic of collusion in public procurement and permanent support.

The author would also like to thank to KSE data analyst Inna Memetova for her guidance with the data and valuable remarks.

I am especially thankful to KSE student Oleksii Romanko for his help with the technical part of the pattern mining algorithm.

Finally, the author is enormously grateful to her family for their colossal belief in me and encouragement to complete this work.

GLOSSARY

ProZorro. Ukrainian electronic public procurement system

AMCU. Antimonopoly Committee of Ukraine

Chapter 1

INTRODUCTION

Public procurement in all countries is a tremendous share of government expenditures each year. As Figure 1 shows, in 2015 it varied from 12.6% to 14.4% in high and low income countries respectively. Particularly in Ukraine, public procurement constitutes a substantial portion of GDP, each year it compiles approximately 13-18%[1]. Government procurement has often been associated with corruption, fraud and collusion, this problem is very spread in many countries. Collusion in procurement leads to the various negative consequences such as artificially overstated prices and low quality of goods and services provided. In Ukraine collusive behavior leads to the annual loss of 10-15% of the state budget[2]. The study about construction firms in Japan finds that collusive activity increases government spending by 3.4 billion USD per year or 0.85% of Japan's GDP (Kawai and Nakabayashi 2015).

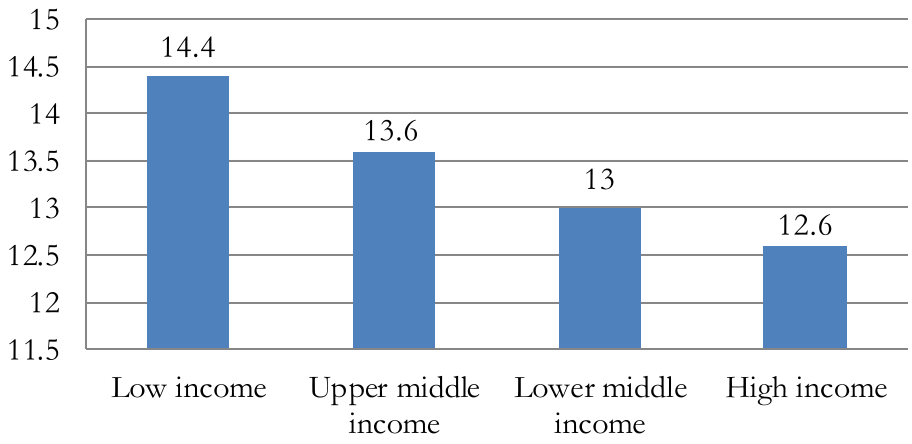

Figure 1. Public procurement as a percentage of GDP in year 2015, by countries' income

---

1 http://www.me.gov.ua/Documents/Detail?lang=uk-UA&id=38c083f3-2571-466a-9583-3b43c2804ad9&title=ReformaDerzhavnikhZakupivel

2 http://www.niss.gov.ua/articles/1414/

The introduction of the new public electronic procurement system "ProZorro" in 2016 obliged all public institutions to conduct the procurement tenders via the electronic system, which enabled to monitor the full process of government procurement. Using "ProZorro" by public enterprises is supposed to increase the transparency and retrenchment of procurement by creating the competitive environment for bidding firms and, therefore, excluding collusion possibility between them. However, firms find various ways of avoiding competition such as unofficial agreement between procurement entity and contractors or collusion among tender participants (which is called "bid rigging"). Thus, the question of the effectiveness of the "ProZorro" system is extremely relevant for policy makers and stakeholders.

Bid rigging schemes can be implemented in various ways such as cover bidding, bid suppression, market allocation and bid rotation. Cover bidding occurs when bidder submits higher bid than that of the predetermined winner, too high bid to be recognized by tendering authority (purchaser) or encloses some unacceptable conditions. Bid suppression implies the agreement between participants, according to which some of the bidders agree to abstain from bidding or cancel previously submitted offer. Market allocation means splitting the market by customers or geographic area between firms. When competitors alternate their turns of being a winner it is a sign of a bid rotation scheme. In my master thesis I intend to focus on the cover bidding, however, these schemes are often implemented in a mixed way.

The main goal of my study is to build a model for assessment of the probability of firm behaving suspiciously in a particular lot. In particular, I investigate how the characteristics of entities and lots are associated with this probability in competitive procurement procedures by using "ProZorro" system data. These factors include particular features of lots and firms, which characterize the auction set-up and output. This research could be valuable for prosecuting authorities in developing a tool for automatic detection of the firms, which behave in suspicious way, to conduct further investigation.

There are a lot of case studies conducted by Ukrainian researchers and the representatives of NGOs, which explore specific suspicious tenders looking at different signs of uncompetitive behavior. Undoubtedly, identifying bid rigging is a complicated process, but there is an exorbitant demand among economists and policy makers in developing a more advanced econometrics-based detection approach. The developed model in my research could be used by Antimonopoly Committee of Ukraine and State Audit Service in order to monitor suspicious tenders and investigate them.

The topic of bid rigging detection was frequently discussed in the literature (Ishii, 2009, Bajari et al, 2001, Porter and Zona, 1992). However, to my knowledge there are no academic papers, which consider developing a model about suspicious behavior of auction participants in the Ukrainian public procurement auctions. Besides this, the methodology that I use is different from the econometric tools applied before.

The methodological approach, which I apply, could be divided into the four parts:

1) running sequential pattern mining algorithm in order to identify firms, which frequently participate in auctions together and one of them wins;

2) identifying firms, which do not change their bid during all rounds of auction;

3) marking those firms, which do not change their bids during all rounds of auction and frequently participate with others as suspicious;

4) estimating econometric model for evaluation of the factors that affect probability of suspicious behavior.

The results of estimations suggest that firms with suspicious behavior in agricultural and food and beverages industries tend to choose the lots with the estimated value, which is slightly above the threshold of 200 thousand UAH or significantly below the threshold value in the case of open competitive procedures. Also in those industries suspicious firms almost never participate in lots only with members of bidding ring. However, in the industry of "Petroleum products, fuel, electricity and other sources of energy" suspicious firms tend to choose below threshold lots with higher estimated values. The entities, which were previously investigated by Antimonopoly Committee of Ukraine tend to possess more suspicious behavior only in Petroleum products, fuel, electricity and other sources of energy industry, while in the other two industries such firms are less suspicious.

The work is structured in the following way: Chapter 2 presents the literature review on the topic of investigating collusive behavior in public procurement auctions; in Chapter 3 there is a methodology applied in current research, Chapter 4 contains the information about the data used. The estimation results are presented in Chapter 5, while Chapter 6 presents the discussion of obtained results and possible policy implications.

Chapter 2

LITERATURE REVIEW

The topic of bid rigging in public procurement auctions was described by numerous theoretical and empirical papers. There are various econometric tests proposed in the literature for bid rigging detection. The researchers make use of several auction characteristics in order to distinguish between collusive and competitive behavior.

In particular, in the study about Japanese construction projects by Kawai and Nakabayashi (2015) the authors use the tests dealing with the bidding strategies. Firstly, authors check the persistence of the identity of the lowest bidder in each round of the first-price sealed-bid auction with rebidding. Second test investigates the optimality of bidding strategy at the second round by looking at the smoothness of the bids' differences distribution. The asymmetric distribution around zero of the differences between the first rounds' three lowest bidders in the second round is a signal of bid rigging. They find evidence of cover bidding patterns that are steady across time, regions and types of works. The similar approach was applied by Haile and Tamer (2003) to the data from the US timber-harvesting auctions. Authors obtain that willingness to pay of each bidder varies as the auction proceeds because of the information inferred from others' behavior.

Many researchers use the analysis of bid price in the collusion detection (Porter and Zona, 1992; Lengwiler and Wolfstetter, 2006). Porter and Zona (1992) look at the differences in behavior of collusive and non-collusive bidders in the sphere of highway construction; in particular, authors analyze the factors, which influence the level of the submitted bid. It is found that the capacity of the firm and job backlog (the number of unfinished jobs) have a positive significant influence on the participants' bid from cartel firms, while for competitive firms these effects are the opposite. Also authors construct a bid ranking of cartel and non-cartel firms and obtain that the bids of competitive bidders and their rank correspond to the firms' incremental costs for taking an additional work. On the contrary, the rank of high bids of cartel members does not coincide with cost measures.

Some authors investigate the influence of reserve prices on the possibility of bidders to arrange a bid rigging scheme. For instance, Thomas (2005) examine how the selection of the reserve price by procuring entity in repeated procurement auction can affect the bidders' ability to act non-competitively. The author develops three strategies of setting a reserve price taking into account the distribution of the bidders' costs and static Nash equilibrium price-setting. The results suggest that cover bidding is more difficult to keep up if the mean of bidders' cost distribution rises. Stepaniuk (2017) considers the influence of reserve price on the competition, which is measured by the number of auction participants, during the dynamic stage of the auction in the Ukrainian public procurement tenders for natural gas, A4 office paper and chicken eggs. The author's hypothesis about the direct relationship between increasing reserve price and improving competition is not supported by the data; however, for tenders for office paper this effect is significant, but considerably low in magnitude.

Recent studies highlight the need for working with the bid price-to-reserve price ratios rather than bid prices or winning bid prices, to avoid the problem of heteroscedasticity. In particular, Bajari and Ye (2001) observe the factors, which influence the bidding behavior of the firms in the highway repair auctions. In order to resolve heteroscedasticity issue, the ratio of the bid price and the estimated price (by engineers) is used. The result of this study is consistent with findings of Porter and Zona (1992). On top of this, the authors propose tests for exchangeability and conditional independence of the firms' simultaneous bids.

Padhi and Mohapatra (2011) using the data on Indian construction projects conduct statistical analysis of bid price-to-reserve price ratios for all bidders and divide the ratios into two significantly different clusters. Ishii (2009), discussed below in more details, suggests that the bid price-to-reserve price ratio, which is higher than 95%, should be a signal of collusion. It is also found that the cluster with higher values of mean and variance of the ratios is associated with collusive behavior, while the other cluster with the low mean and variance corresponds to competitive bidding. In addition to this, he constructs the graphical tool, which helps to detect the incidence of collusion right after the opening of the price bids in the auction. Imhof (2017) and Abrantes-Metz et al. (2006) argue that the variance of bids declines in the case of collusion.

Ishii (2009) using the public procurement data on compensation consulting works constructs the index, which indicates whether bidder won or lost in an auction and the score of each bidder, which depends on the net balance of favors between each two bidders in the ring. He examines how the fact of winning the auction depends on the score between bidders and other control variables. The results seem to confirm that the bidders exchange favors to be the winner, because the winners are more likely to have a positive score against losers. Also it is shown that the losers participate in auctions more frequently than the winners and the winners operate longer in the market.

Some authors consider a possible communication between bidders as a sign of bid rigging. Using laboratory experiments, Agranov and Yariv (2016) study how the number of interactions (communication or transfers) available to bidders influence the price in one-shot first- and second-price sealed-bid auctions. They find that communication alone is associated with the rare bid rigging cases and causes significant, but minor decrease in price, while the cases of communication with transfers coincide with the common documented bid rigging.

To sum up, the topic of uncompetitive behavior of firms was investigated from the several prospects and definition of collusion varies from paper to paper. Frequently authors concentrate on the bidding patterns of auction participants and legal case studies about convicted firms and argue that high relative bids signal about cover bidding. Some studies investigate the relationship between bids of several bidders itself. There are a lot of studies about the bid rigging in the US public procurement auctions with econometric models, however, there are no such studies about Ukrainian public auctions. The possible reason of that is the lack of sufficient legal data about the convicted firms in Ukraine. This thesis is aimed to solve the problem of absence of the legal information by developing own indicator of suspicious behavior using the available data.

Chapter 3

METHODOLOGY

According to the Ukrainian Law "On Public Procurement", the competitive procedure via the electronic system is obligatory for procurement of goods with the value over 200 thousands UAH for all spheres except the sphere of gas, energy, water supply and others, where this threshold is one million UAH. However, if the procuring entity wants to hold a competitive tender via the "ProZorro" system even for lower than threshold amount, there are no restrictions on using the electronic system.

Ukrainian competitive procurement auctions are dynamic and held in four stages (rounds). The first or so-called zero round is similar to a sealed-bid auction, where participants do not observe the bids of other participants. At this round the order of bidding at the next round is decided: participant, who claimed the highest bid, goes first at the second round, thus, bidder with the smallest price has an advantage of acquiring information about the others' bids and choose the most beneficial bid. The same scheme is applied to every round. The other three rounds are held in the format of dynamic English auction, where the information about the number of participants and their bids is observed by all bidders.

The key variable in this analysis is the probability of a firm being suspected in uncompetitive behavior (being a part of a bidding ring) in the particular lot given the characteristic of the firm (indicator of being prosecuted by Antimonopoly Committee of Ukraine). Uncompetitive behavior in this work is identified in the following way: firm participates frequently with the other firms in auctions and one of them wins and firm does not change its bid during all rounds of auction. Further in this work these firms would be called "suspicious".

The first step in my analysis is to identify firms, who frequently participate together in auctions. With this aim the sequential pattern mining algorithm is used, which is introduced in the paper of Ayres et al. (2002). This particular algorithm is extremely efficient, when the database has very long sequential patterns, which is suitable for database in this work.

The methodology of finding sequential patterns (Ayres et al, 2002) is the following: let $I = \{i_1, i_2, \dots i_n\}$ be a set of items (firms, tenderers). The subset $X \subseteq I$ is called an itemset (set of tenderers in each lot) and $|X|$ is the size of X. A sequence $s = (s_1, s_2, \dots s_m)$ is an ordered list of itemsets and $s_i \subseteq$ I , i $\in \{1, \dots m\}$. The size, m, of a sequence is the number of itemsets in the sequence (the sequence of lots, at which particular tenderers participate). The length l of a sequence $s = (s_1, s_2, \dots s_m)$ is defined as $l \stackrel{\text{def}}{=} \sum_{i=1}^{m} |s_i|$. A sequence $s_a = (a_1, a_2, \dots a_n)$ is contained in another sequence $s_b = (b_1, b_2, \dots b_m)$ if there exist integers $1 \leq i_1 < i_2 < \dots < i_n \leq m$ such that $a_1 \subseteq b_{i1}, a_2 \subseteq b_{i2}, \dots, a_n \subseteq b_{in}$. If sequence s_a is contained in sequence s_b, then s_a is called a subsequence of s_b and s_b is a supersequence of s_a.

The database of our interest is a set of tuples (sequence, lot id, X, winner), where sequence is the number of sequence, lot id is an identifier of a lot and X is an itemset of tenderers such that $X \subseteq I$, winner is an id of a winning firm. The structure of the database is displayed in Table 1.

In addition to the condition that firms participate together in the auction with each other, one of them should necessarily win. In the work of Bajari and Ye (2001) about public-sector seal coat contracts authors consider the number of simultaneous bids, which is more than 4 out of 495 or 0.8% of auctions in the data set. Following this paper in using pattern mining algorithm in the analysis of public procurement data, as the threshold of frequency (length of a sequence) I assume 0.8% as minimum of all lots in the sample, which corresponds to the 45 lots.

Table 1. Dataset sorted by sequence number

Sequence Number	Lot ID	Itemset (set of tenderers)	Winner
1	1	{a,b,c}	a
1	2	{a,b,c}	b
1	3	{a,b}	a
2	4	{c,d,e}	c
2	5	{c,e}	c

Among those firms that were identified by the algorithm, only firms that did not change their bid during the auction are denoted as 'suspicious'. The lot participants can change their bid by extremely minor value in order to rebid the other participants and it is consistent with the competitive behavior. This is the case of Ukrainian open public auction procedures due to the auctions' design.

Next, we estimate how various factors (characteristics of lot and bidders) explain the probability of firm being suspected in uncompetitive behavior. Similar approach is used, for example, in credit score models to evaluate probability of default of economics agents conditional on a set of individual characteristics. In particular, fraud detection models use the data from financial statements with the aim of identification fraudulent cases of deliberate financial statements falsification (Jan 2018).

The second stage of this model considerably increases the advantages of the fraud detection model, because it allows to incorporate external factors that can explain suspicious behavior of the firms. Those factors can increase efficiency of the model in two ways. Firstly, it can sufficiently increase the prediction accuracy of the model, utilizing predictive power of external factors. Secondly, it enables usage of the model on the new, previously unknown data, in order to score the events, which are close to the already studied ones.

The logistic regression model, as a model of choice in the current work, was proved to be one of the most robust and accurate in terms of predictive power for the scoring purpose. For example, Gurny and Gurny (2013) compare the modeling approaches for the probability of default estimation in the banking industry. Author finds that logistic model tends to be the best model among compared, as well as the model needs less explanatory variables to achieve competitive results.

We follow similar approach to the probability of default scoring. In the same manner, as the probability of default scoring model estimates the probability of bad event occurrence (default), current model measures the probability of firm to be suspected in cover bidding.

The one of the main benefits of probability modeling is ability to determine the threshold, which determines the classification behavior of the model. Depending on the financial costs of the wrong decision (detect competitive firms as suspicious ones), threshold can be modified for the individual purpose in order to account for the costs of error.

We model the probability of being suspected in uncompetitive behavior as a function of lot and participants' characteristics. This model is constructed in order to provide a score, which is related to the probability that entity will carry out equivocal behavior in the lot with certain features. To estimate this model the logistic regression is used:

Pr(suspicious) = f(***Lot characteristics*** (qty_part, qty_complaints, lnestim, city_same, Difference btw 1st 2nd rel pr, above_threshold, lnestim*above_threshold, above_threshold*qty_complaints, above_threshold*`Difference btw 1st 2nd rel pr`),
Participant's characteristic (amkuflag)),

where

- suspicious - dependent binary variable is equal to 1 if firm participated in auctions together with the same firms more than in 0.8% of all auctions in the data and did not change the bid during four rounds and 0 otherwise.

Lot characteristics:

- qty_part – number of participants in a lot;
- qty_complaints – amount of complaints in a lot (the complaint of a bidder about a tendering authority's decision typically occurs when the selection of the winner is not transparent);
- lnestim – estimated value of the lot (reserve price) in logs;
- city_same – dummy (=1 if at least two bidders registered at the same city/town, 0 otherwise);
- Difference btw 1st 2nd rel pr – difference between the winning bid and second lowest bid at the third round relative to the estimated value (in percentage terms);
- Above_threshold – dummy (equal to 1 if the estimated value of a lot is above the 200 000 UAH threshold, 0 otherwise);
- lnestim* above_threshold – interaction term of the estimated value in log and the threshold dummy;
- above_threshold*qty_complaints – interaction term of the amount of complaints and the threshold dummy;
- above_threshold* `Difference btw 1st 2nd rel pr` – interaction term of the relative difference between the winning bid and second lowest bid at the third round and the threshold dummy.

Participant's characteristic:

- amkuflag – dummy for Antimonopoly Committee of Ukraine (AMCU) prosecution (=1 if firm was on the AMCU list, 0 otherwise).

Number of participants is expected to have a positive effect, because it is easier for firms to hide a bid rigging schemes if there are more than two participants in the auction.

The relationship between the number of complaints and suspicious behavior of firms should be positive as usually firms or civic activists complain if they think that the selection of the winner was not fair, clear and transparent.

The auction reserve price (estimated value) is also supposed to be positively related to the probability of being suspected in cover bidding, because it attracts firms for the possibility of gaining high profit and thus, stimulates creation of 'cartels' in order to grab the highest possible price. The estimated value is taken in log in order to normalize it, because its distribution is left-skewed.

Lots with firms with the same city code are also positively related to cover bidding, because frequently firms register another "pocket" firm at the same area or address for auction to take place.

The difference between the first and second relative bid is expected to have positive effect, because it captures the intensity of competition: the gap between the winner and runner-up in case of competitive behavior should be small. Relative measures were used instead of absolute values in order to prevent the potential problem of heteroscedasticity (Ishii, 2009). The emphasis is made on the third round, because they are the most crucial: the winner is determined at this round.

The fact the firm was prosecuted by AMCU makes it more suspicious as it was already investigated for violation of competitiveness.

The threshold dummy was included into the model with the aim of capturing the difference between the lots, to which open competitive procedure is applied mandatory or voluntary. On top of this, in this way we can distinguish between the lots with high and low estimated values.

The interaction of the estimated value in log and the threshold dummy was chosen in order to observe suspiciousness of firms at the lots with the estimated values above the threshold as higher estimated values (above the 200 000 UAH) mean bigger amount of profits, which are more attractive for collusive bidders.

The interaction of the amount of complaints and the threshold dummy was included into the model with the aim of capturing the difference between amount of complaints in the lots below and above the threshold values. As filing complaint is expensive (bidder has to pay 5000 UAH to file a complaint with the AMCU[3] in addition to the payment for lawyers services), complaining on the winner selection makes bigger sense for participants in the lots with the more valuable lots.

Interaction of the relative difference between the winning bid and second lowest bid at the third round and the threshold dummy is supposed to have a positive sign. Lots with the reserve price, which is substantially bigger than the threshold value, are very attractive to the potential members of bidding rings. However, the coefficient on this interaction should be higher in magnitude in comparison with the relative difference between the winning bid and second lowest bid itself, because there is frequently the case, when the estimated value of a lot is established on the basis of the full budget allocated instead of engineer or market analyst estimates, because government organizations are not very much concerned about the resource spending efficiency or economy. Thus, bidders have a possibility to decrease their bids in each round by the substantial percent (if their bid is significantly higher than their valuation of the goods).

[3] http://www.amc.gov.ua/amku/control/main/uk/publish/article/87468

However, each industry of interest has specific features: that is why the model estimation is done for each of them separately. The markets of Agricultural, farm products, fishery products, forestry and related products and Food, beverages, tobacco and related products are highly competitive; however, the goods traded might have slightly different specifications. While the Petroleum products, fuel, electricity and other sources of energy industry in Ukraine is characterized by the homogeneous products, it is a form of oligopolistic competition and Antimonopoly Committee of Ukraine dedicates a lot of its effort to collusion in this sphere.[4]

In order to estimate the accuracy of model predictions, the receiver operating characteristic (ROC) curve is used as it is a widely used tool in modern statistics. By choosing the most suitable threshold value for the score of suspicious behavior, we can construct the graphical representation of the true and false positive rates and compute the area under the curve (AUC) score.

[4] http://www.amc.gov.ua/amku/control/main/uk/publish/article/130459

C h a p t e r 4

DATA DESCRIPTION

The main data source, which is used, is electronic public procurement system "ProZorro", which includes the information on the competitive procedures in Ukraine across all industries held in the system during year 2016.

This work is focused on the following industries: Agricultural, farm products, fishery products, forestry and related products; Petroleum products, fuel, electricity and other sources of energy; Food, beverages, tobacco and related products, because the goods, which are traded are quite homogeneous and there are no market entry barriers, thus, the quantity of competitors should be enough for excluding the possibility of few suppliers at all.

The data is structured by firms and lots; it contains the information about tendering authority (organizer), bidders (participants, tenderers): their address, city, ZIP codes; estimated value for each lot (reserve price), bid of each tenderer at all rounds, the information about goods or works supplied (name and industry, to which it belongs). Overall, my dataset has approximately ten thousands observations.

The second data source, which was processed, is the data of Antimonopoly Committee of Ukraine (AMCU) on the firms, which were prosecuted for uncompetitive behavior. This data covers the cases of competition violation investigation of the period up to year 2016. There are 696 cases out of 10 thousand, when those firms took part in auctions.

The dependent variable in this analysis is a dummy variable for suspicious firm constructed based on the results of the sequential pattern mining algorithm and constant bid indicator. Petroleum products, fuel, electricity and other sources of energy industry has the highest share of tenderers, which participate frequently with other participants detected by the sequential pattern mining algorithm (Table 2). In Petroleum products, fuel, electricity and other sources of energy and Food, beverages, tobacco and related products industries approximately 20% of entities do not change the bid during the four stages of the auction. The Petroleum products, fuel, electricity and other sources of energy industry has the highest number of suspicious firms participating in auction.

Table 2. Summary statistics of dummy variables across divisions, % of all lots

	Prosecuted by AMCU	Same city	Frequent participant	Constant bid	Suspicious firms
03000000-1 Agricultural, farm products, fishery products, forestry and related products	0.5	3.3	0.2	4.9	0.2
09000000-3 Petroleum products, fuel, electricity and other sources of energy	3.9	10.5	10.7	23.2	7.0
15000000-8 Food, beverages, tobacco and related products	1.2	13.9	6.2	21.9	5.2

As Table 3 shows there are 12.4% of cases, when suspicious firms took part in an auction in the dataset.

Table 3. Summary statistics of dependent variable

	Min	Max	Mean	Median	St. dev.
Suspicious firm	0	1	0.124	0	0.33

There are 7267 cases out of all sample, when the bid of participant was not changed during the four rounds, 2482 times when frequent simultaneous bids take place in auctions and 1808 times when it happens jointly. According to the Table 2, Petroleum products, fuel, electricity and other sources of energy industry have the largest number of cases when firms have constant bids during all auction process and have a high percentage of firms, which participate together frequently in procurement procedures. Also the Food, beverages, tobacco and related products industry has considerable percentage of suspicious tenderers.

According to the Table 4, there are on average three bidders in an auction, the average difference between the first and second relative prices in the third round of the auction is 0.05. There are 10% of firms, which were prosecuted by the AMCU. The complaining does not occur at up to 50% of all lots (the median equals to zero), which is expected, because the filing a complaint is not free of charge, it costs 5000 UAH in case of appeal of the procedure for procurement of goods or services and 15000 UAH in case of procurement of works.[5] 27% of firms have identical city codes, which means that they are registered at the same city or town. There are 73% of lots, which have the estimated value, which is higher than the threshold of 200 thousand UAH.

The descriptive statistics for the Agricultural, farm products, fishery products, forestry and related products industry is shown at the Table 4. The range of the reserve prices of lots is quite high; however, as median for the estimated value is much bigger than 200 thousand UAH, one can see that the amount of above threshold lots prevails in the dataset, in particular, it is 68%. There are on average 3 participants in the lot. The firms previously caught by the Antimonopoly Committee of Ukraine took part in the 6% of lots. In the 69% of lots there was a case, when entities come from the same city. The difference of the first and second relative bids is on average 10%. There are 9% of cases, when suspicious firm was one of the participants in the lot.

[5] http://www.amc.gov.ua/amku/control/main/uk/publish/article/87468

Table 4. Summary statistics of the Agricultural, farm products, fishery products, forestry and related products industry

	Min	Max	Mean	Median	St. dev.
Estimated value, thous. UAH	6.60	3731	535	323	690
Number of participants	2.00	9.00	3.32	3.00	1.51
AMCU prosecution	0.00	1.00	0.06	0.00	0.23
Amount of complaints	0.00	1.00	0.01	0.00	0.08
Same city/town	0.00	1.00	0.69	1.00	0.46
Difference btw 1st 2nd rel pr	0.00	92.00	9.81	2.86	14.02
Suspicious	0.00	1.00	0.09	0.00	0.16
Above threshold	0.00	1.00	0.68	1.00	0.46
Number of observations = 1080					

The Table 5 reveals that in the Petroleum products, fuel, electricity and other sources of energy industry there are 11% of cases of suspicious entities involving into the auction. The reserve prices of lots in this sphere are much higher, than in the sphere of Agricultural, farm products, fishery products, forestry and related products, that is, the motivation to create bidding rings increases. However, the amount of bidders, which have the same city or town of registration, is smaller than in the sphere of agricultural products. There are about 84% of above threshold lots, which is much higher than in the spheres of Agricultural, farm products, fishery products, forestry and related products and Food, beverages, tobacco and related products

The mean estimated value of all lots in Food, beverages, tobacco and related products industry is 506 thousand UAH (Table 6), which is quite similar to the Agricultural, farm products, fishery products, forestry and related products, the average number of participants in the lot is also three. There are approximately 70% of cases, when firms registered at the same city took part in the auctions.

Table 5. Summary statistics of the Petroleum products, fuel, electricity and other sources of energy industry

	Min	Max	Mean	Median	St. dev.
Estimated value, thous. UAH	2.34	15100	1241	800	1219
Number of participants	2.00	11.00	3.49	3.00	1.63
AMCU prosecution	0.00	1.00	0.11	0.00	0.31
Amount of complaints	0.00	1.00	0.03	0.00	0.17
Same city/town	0.00	1.00	0.37	0.00	0.48
Difference btw 1st 2nd rel pr	0.00	97.80	4.61	2.24	7.23
Suspicious	0.00	1.00	0.11	0.00	0.32
Above threshold	0.00	1.00	0.84	1.00	0.37
Number of observations = 4392					

Table 6. Summary statistics of the Food, beverages, tobacco and related products industry

	Min	Max	Mean	Median	St. dev.
Estimated value, thous. UAH	4.83	3684	506	303	642
Number of participants	2.00	10.00	3.41	3.00	1.36
AMCU prosecution	0.00	1.00	0.03	0.00	0.17
Amount of complaints	0.00	1.00	0.02	0.00	0.15
Same city/town	0.00	1.00	0.70	1.00	0.46
Difference btw 1st 2nd rel pr	0.00	99.96	6.88	1.73	10.45
Suspicious	0.00	1.00	0.17	0.00	0.37
Above threshold	0.00	1.00	0.65	1.00	0.48
Number of observations = 4859					

Chapter 5

EMPIRICAL RESULTS

The logit models for three industries were estimated and the results are presented in the Table 8. The dataset was divided into in-sample and out-of-sample subsets.

The ROC curves and computed accuracies suggest that models do quite well in predicting the dependent variable (Table 7). The models for "Agricultural, farm products, fishery products, forestry and related products" and "Food, beverages, tobacco and related products" industries show extremely good accuracies. However, the estimation of "Petroleum products, fuel, electricity and other sources of energy" might need some improvement because of the issues, which are discussed later.

Table 7. Accuracy of model predictions

	Pseudo R squared	Accuracy (AUC)
Agricultural, farm products, fishery products, forestry and related products	0.48	0.98
Petroleum products, fuel, electricity and other sources of energy	0.35	0.89
Food, beverages, tobacco and related products	0.57	0.94

Analysis for "Agricultural, farm products, fishery products, forestry and related products" industry reveals that amount of complaints has expected sign on the probability of firm being suspicious. The presence of one complaint in a lot is associated with the increase in the probability of suspicious behavior by 20%, which is significant.

However, number of participants has significant negative impact, which suggests that agricultural industry is highly competitive and firms, which are engaged in collusive schemes, often participate in auctions only with bid rigging counterparts. It is quite difficult for collusive firms to hide their suspicious behavior in such "heterogeneous" lot, because if another tenderer, which is not part of the bidding rings, suspects unfair selection of winner, he will submit a complaint to the Antimonopoly Committee of Ukraine. This result is also supported by the negative coefficient on the reserve price: high reserve price attracts more participants and public attention to the lot, thus, there is a high probability that uncompetitive behavior will be detected and prosecuted. As a result, collusive firms prefer lots with smaller potential profit.

Also, the negative impact of difference between first and second relative price at the third round in above threshold lots reveals that there is a slight difference between the winner bid and the second bid, thus, suspicious firms almost always participate in lots with the non-collusive firms. On top of this, if entities are registered at the same city, it lowers the probability of being suspicious, however, the magnitude is quite small (6.5%). The coefficient of AMCU prosecution is negative and rather small, but very significant. The possible reason of that could be that AMCU pays more attention to the entities, which were already prosecuted, thus, they do not have an incentive to distort competition at the competitive procurement procedures as they can be easily caught on cheating. However, the insignificant coefficient on the interaction of reserve price and threshold value and significant (but extremely small) coefficient on the interaction of above threshold dummy and relative difference reveals that, in general, there is no huge difference between the below- and above-threshold value lots for the products in this category.

Estimation results for the tenders for "Petroleum products, fuel, electricity and other sources of energy" reveal that the number of participants and same registration city of participants are not associated with a higher or lower probability of being suspicious.

Table 8. Estimation results of logit model (marginal effects)

Variables	Agricultural, farm products, fishery products, forestry and related products	Petroleum products, fuel, electricity and other sources of energy	Food, beverages, tobacco and related products
Estimated value in log	-0.030* (0.014)	0.020* (0.012)	-0.032*** (0.007)
Number of participants	-0.026** (0.009)	0.003 (0.003)	-0.031*** (0.005)
Dummy for AMCU prosecution	-0.029*** (0.004)	0.140*** (0.020)	-0.170*** (0.003)
Same city of participants	-0.065*** (0.011)	-0.015 (0.011)	0.020* (0.010)
Amount of complaints	0.203*** (0.046)	-0.135*** (0.005)	-0.171*** (0.003)
Above threshold	0.038 (0.161)	0.185*** (0.046)	-0.369 (0.252)
Difference btw 1st 2nd rel pr	0.016 (0.036)	-0.002 (0.147)	-0.048*** (0.009)
Above threshold * estimated value in log	0.000 (0.014)	-0.030* (0.013)	0.005 (0.010)
Above threshold * Amount of complaints		0.862*** (0.005)	0.349*** (0.061)
Above threshold * Difference btw 1st 2nd rel pr	-0.002* (0.078)	0.002 (0.159)	0.039*** (0.010)

Notes: Standard errors in parentheses. * if p-value < 0.05, ** if p-value < 0.01, *** p < 0.001.

As expected higher estimated value for lots below the threshold is associated with substantially higher probability of suspicious behavior: 10 percent increase in estimated value adds almost 20 percent to this probability. On the contrary, the negative coefficient on interaction term of estimated value and threshold dummy reveals that suspicious firms tend to choose lots with the estimated value, which is slightly above the threshold value: the 10% rise in estimated value is connected with the 10% decline in probability of suspiciousness. The feature of "Petroleum products, fuel, electricity and other sources of energy industry" is that the vast majority of procurement here is conducted on the values that are higher than 200 thousand UAH.

Already prosecuted by AMCU entities tend to behave by 14% more suspiciously than those firms, which were not caught.

The presence of one complaint in a lot with reserve price below the threshold value is associated with the decrease in the probability of firms behaving suspiciously by 13.5%. However, at the above threshold lots one complaint is linked with 73% of firm's suspiciousness, which is considerably high.

The observed relationships in "Food, beverages, tobacco and related products" industry indicate that suspicious firms are not attracted by the high reserve price in this industry in below threshold lots, which is similar to the Agricultural industry. However, as the interaction term of the threshold dummy and the difference between the first and second relative bid at the third round is positive and significant, it means that increase in the difference between the first and second relative bid in the lot with the reserve price above the 200 thousand UAH by 10% is associated with 39% higher probability of suspicious behavior of the bidder, than in lots with the estimated value under the threshold, which is tremendously high. At the same time, the overall effect of growth of relative difference is negative for lots above the threshold.

The coefficient on the amount of complaints for below threshold lots is negative, which suggests that collusive firms are unlikely to be the only participants in a lot (with competitive ones). However, for lots with the reserve price above the threshold, this effect is positive: the presence of complaint the probability of suspicious behavior by 17.8%.

The significant negative coefficient on the number of participants is associated with the lower probability of suspicious behavior, which is consistent with the finding about amount of complaints. The negative coefficients on the prosecuted firms by AMCU suggest that those firms do not possess risk of behaving suspiciously no more, such firms are, on average, by 17% less suspicious.

In order to check the robustness of the estimated results, in the sequential pattern mining algorithm I changed the threshold for the detection of bidders frequently participating together in tenders from 0.8% to 0.4% of the total amount of lots. By decreasing the threshold, I relax the selection condition for the algorithm. Hence, if the direction and magnitudes of coefficients remain the same, the results of my estimation model are robust. This is actually confirmed in Table 9 (in Appendix).

We should be cautious in interpreting coefficients on amount of complaints and AMCU prosecution due to the issues of limited data. As it was mentioned before, the complaint application to the AMCU costs 5000 UAH in case of goods procurement, which is fixed for all lots regardless of industry and estimated value. Thus, AMCU accepts a few complaints about suspicious cases from participants of lots with low estimated value and many applications from high estimated value lot participants. On top of this, the capacity of Antimonopoly Committee of Ukraine is not sufficient to deal with all the complaint applications in timely manner[6]; that is why a lot of investigations are conducted with lag and the data on AMCU prosecution is available only up to year 2016. Also, the problem with AMCU data is that there is no unique data on the cases, which were identified by AMCU as

[6] http://cep.kse.org.ua/article/choho-varta-ideya-pominyaty-tsinu-skarhy-v-AMKU/brief.html

collusion (actual data contains only investigated firms without clear judicial conclusions).

The accuracy of the prediction of the model on the data about "Petroleum products, fuel, electricity and other sources of energy" reveals that this model might need some refinement, because the data available in the electronic public procurement system is not sufficient, because the suspicious behavior of firms in this industry depends not only on the factors, which are present in our model.

In addition to this, the overall data quality in the electronic public procurement system "ProZorro" needs some improvement, because there is a lot of missing observations about the firms' cities of registration and ZIP codes.

In addition, there is an issue about the threshold of the sequence of lots, which is considered to be frequent. There is no literature about using pattern mining algorithm in investigation of Ukrainian procurement auctions, thus, further research is needed.

C h a p t e r 6

CONCLUSIONS

The purpose of this research is to investigate the factors, which are associated with the probability of firm behaving suspiciously given specific firm and lot characteristics in three industries: Agricultural, farm products, fishery products, forestry and related products; Petroleum products, fuel, electricity and other sources of energy; Food, beverages, tobacco and related products. The sequential pattern mining algorithm and constant bid behavior indicator is used in order to identify suspicious tenderers and logistic regression estimation is applied to evaluate factors related to probability of being suspicious.

The results suggest that lots with high estimated value do not seem to attract suspicious firms in agricultural and food and beverages industries in lots with both below and above threshold estimated values. Typically, collusive entities tend to participate in the lots with reserve price, which is slightly above average or significantly below the threshold value. Also in these industries suspicious firms almost never participate in lots only with members of bidding ring. However, in the industry of "Petroleum products, fuel, electricity and other sources of energy" suspicious firms tend to choose below threshold lots with higher reserve prices. However, in all three industries suspicious tenderers have relatively high bids at the final auction round.

The firms, which were previously investigated by Antimonopoly Committee of Ukraine, are more suspicious only in Petroleum products, fuel, electricity and other sources of energy industry. In the other two sectors the effect is opposite, which implies that such prosecuted firms are less suspicious. The possible issue here is that only a few amount of indeed collusive firms was taken into account of AMCU and lots, at which these firms participate, have zero quantity of complaints. This issue arises from the fixed price for complaining to the AMCU for lots with different reserve prices. Usually complaining make sense only if the lot estimated value is significantly higher that it's' reserve price. Thus, the price of filing a complaint should be revised. On top of this, the capacity of AMCU does not allow it to proceed all uncompetitive cases, thus, there is a need to increase it.

The model developed can be used as a collusion detection mechanism in order to identify cover bidders in public procurement auctions. The extreme importance of the model developed is a combined mining-and-regression nature of the final model, which is different to the models that are commonly used in fraud detection sphere. It allows not only to detect patterns, but also gives possibility for automation of the prosecution process, because it allows to incorporate previous decisions of the Antimonopoly Committee of Ukraine into the model, with the help of the regression model at the second stage of the detection process. In such manner, previous decisions of the AMCU regarding the result of bidding investigation might be used to determine the probability of fraud in the similar cases to the already investigated. There are more factors that might be connected to the suspicious behavior of the entity and hence be incorporated in the model: e.g. the date of the firm registration (if the entity was registered in two month before the auction, it might be the case of creating a "pocket" firm by another bidder), entity owner and others.

During almost two years since the launch and obligatory use of electronic public procurement system in Ukraine collusive firms adapted to the approaches and mechanisms used by the Antimonopoly Committee of Ukraine and civic organizations, which try to monitor and detect improper practices (including bid rigging) in public procurement. Usually, the biggest attention and loudest scandals are concentrated on tenders either with high-valued lots (over one million UAH) or cases when procuring entity purchases a good, which is extremely unusual for this particular buyer, for a price completely different from the average in the market.

The possible policy implication for the Antimonopoly Committee of Ukraine and civic organizations is to focus on the lots, which are not remarkable in terms of the estimated value, because suspicious firms tend to avoid these auctions as they attract huge portion of public attention and it creates some inconveniences for them in implementing the bid rigging schemes. While at the same time, the lots with moderate estimated value seem to be less attractive for investigation.

Also there is an implication for the Ministry of Economic Development and Trade to consider the engineers or market analyst estimates and develop a guidance about the establishing a reserve price for procuring entities, because often those prices are overestimated.

WORKS CITED

Abrantes-Metz, Rosa M., Luke M. Froeb, John Geweke, and Christopher T. Taylor. 2006. "A variance screen for collusion." *International Journal of Industrial Organization* 24, no. 3: 467-486.

Agranov, Marina, and Leeat Yariv. 2016. "Collusion through communication in auctions." *Games and Economic Behavior,* no. 107: 93-108.

Ayres, Jay, Jason Flannick, Johannes Gehrke, and Tomi Yiu. 2002. "Sequential pattern mining using a bitmap representation." *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, July 23-26, 2002, Edmonton, Alberta, Canada.

Gurný, Petr, and Martin Gurný. 2013. "Comparison of Credit Scoring models on probability of default estimation for US Banks." *Prague Economic Papers*, no.2: 163-180.

Bajari, Patrick, and Lixin Ye. 2003. "Deciding between competition and collusion." *Review of Economics and statistics* 85, no. 4: 971-989.

Haile, Philip A., and Elie Tamer. 2003. "Inference with an incomplete model of English auctions." *Journal of Political Economy* 111, no. 1: 1-51.

Imhof, David. 2017. "Simple statistical screens to detect bid rigging." Working papers SES, no. 484, Université de Fribourg.

Ishii, Rieko. 2009. "Favor exchange in collusion: Empirical study of repeated procurement auctions in Japan." *International Journal of Industrial Organization* 27, no. 2: 137-144.

Jan, Chyan-long. 2018. "An Effective Financial Statements Fraud Detection Model for the Sustainable Development of Financial Markets: Evidence from Taiwan." *Sustainability* 10, no. 2: 513.

Kawai, Kei, and Jun Nakabayashi. 2014. "Detecting large-scale collusion in procurement auctions." Working paper, UC Berkeley, Berkeley, CA.

Lengwiler, Yvan, and Elmar Wolfstetter. 2006. "Corruption in procurement auctions." *Handbook of Procurement*, Ch.16: 412-429.

Padhi, Sidhartha S., and Pratap KJ Mohapatra. 2011. "Detection of collusion in government procurement auctions." *Journal of Purchasing and Supply Management* 17, no. 4: 207-221.

Stepaniuk, Oleksa. 2017. "Reserve price above the market level does not attract more participants to public procurement auctions. Evidence from Ukraine." VoxUkraine, June 18, 2017. https://voxukraine.org/uk/prozorro-yak-zavishhena-ochikuvana-vartist-pogirshuye-rezultat-auktsionu-ta-chi-treba-z-tsim-borotisya.

Thomas, Charles J. 2005. "Using reserve prices to deter collusion in procurement competition." *The Journal of Industrial Economics* 53, no. 3: 301-326.

APPENDIX A

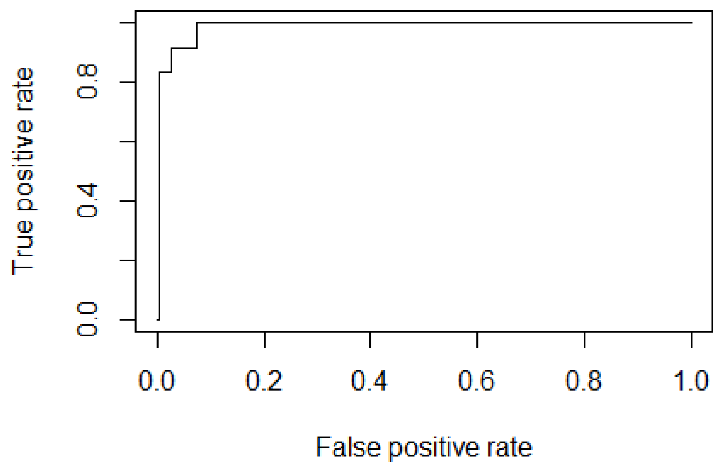

Figure 2. ROC Curve (Agricultural, farm products, fishery products, forestry and related products)

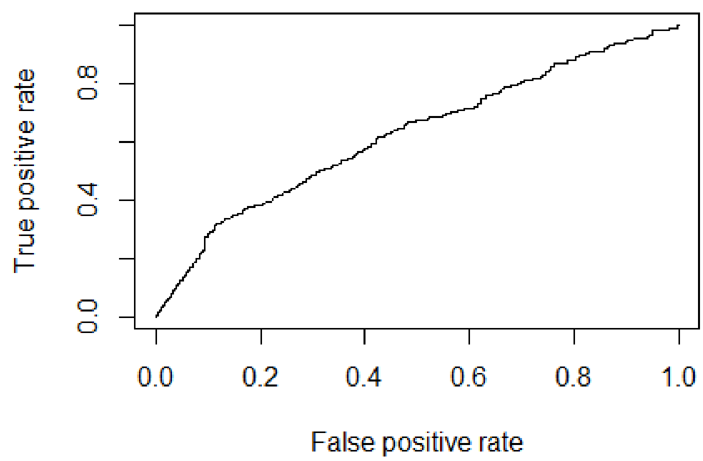

Figure 3. ROC Curve (Petroleum products, fuel, electricity and other sources of energy)

Appendix A continued

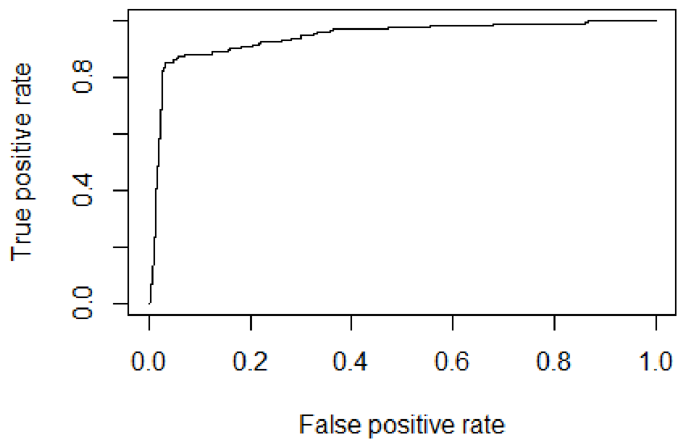

Figure 4. ROC Curve (Food, beverages, tobacco and related products)

APPENDIX B

Table 9. Estimation results of logit model (marginal effects) with 0.4% threshold

Variables	Agricultural, farm products, fishery products, forestry and related products	Petroleum products, fuel, electricity and other sources of energy	Food, beverages, tobacco and related products
Reserve price in log	-0.030* (0.014)	0.019* (0.011)	-0.032*** (0.006)
Number of participants	-0.026** (0.009)	0.003 (0.003)	-0.030*** (0.005)
Dummy for AMCU prosecution	-0.029*** (0.004)	0.139*** (0.020)	-0.169*** (0.003)
Same city of participants	-0.065*** (0.011)	-0.014 (0.011)	0.020* (0.010)
Amount of complaints	0.203*** (0.046)	-0.135*** (0.004)	-0.171*** (0.003)
Above threshold	0.038 (0.161)	0.184*** (0.047)	-0.368 (0.251)
Difference btw 1st 2nd rel pr	0.016 (0.036)	-0.001 (0.001)	-0.004*** (0.001)
Above threshold * Reserve price in log	0.000 (0.014)	-0.030* (0.013)	0.005 (0.010)
Above threshold * Amount of complaints		0.862*** (0.004)	0.349*** (0.060)
Above threshold * Difference btw 1st 2nd rel pr	-0.002* (0.078)	0.001 (0.001)	0.003*** (0.001)

Notes: Standard errors in parentheses. * if p-value < 0.05, ** if p-value < 0.01, *** p < 0.001.