Identifying Bid Leakage In Procurement Auctions: Machine Learning Approach

DMITRY IVANOV* and ALEXANDER NESTEROV*, Higher School of Economics, Russia

We propose a novel machine-learning-based approach to detect bid leakage in first-price sealed-bid auctions. We extract and analyze the data on more than 1.4 million Russian procurement auctions between 2014 and 2018. As bid leakage in each particular auction is tacit, the direct classification is impossible. Instead, we reduce the problem of bid leakage detection to Positive-Unlabeled Classification. The key idea is to regard the losing participants as fair and the winners as possibly corrupted. This allows us to estimate the prior probability of bid leakage in the sample as 16%, as well as the posterior probability of bid leakage for each specific auction.

ACM Reference Format:
Dmitry Ivanov and Alexander Nesterov. 2019. Identifying Bid Leakage In Procurement Auctions: Machine Learning Approach. In *ACM EC '19: ACM Conference on Economics and Computation (EC '19), June 24–28, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3328526.3329642

We study "requests for quotations" – small and frequent online first price sealed-bid procurement auctions. These auctions can suffer from **bid leakage** – the corruption scheme where procurer illegally provides his favored participant with the information about the bids of the other participants. Our goal is to estimate how widespread bid leakage is in general, as well as how likely it is that each particular auction is corrupted. To this end, we analyze the dataset of more than 600 000 Russian requests for quotations that took place from January 2014 to March 2018.

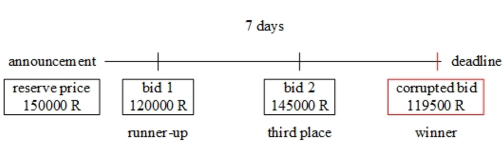

Fig. 1. Request for quotations with leaked bids

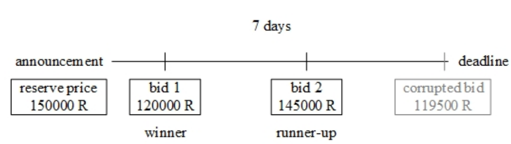

Fig. 2. Placebo auction example

Our work is inspired by [1] who observed the patterns that are likely to reflect rational behavior of the favoured participant that received leaked bids (see Figure 1). These participants are: bidding last (1) and close to the deadline (2) to ensure they know all other bids; undercutting and hence winning by a small margin (3) to maximize their profit.

We use these three patterns to determine whether a particular auction has been corrupted by bid leakage. Our approach is to reduce the problem to Positive-Unlabeled Classification by considering the runner-ups as fair participants and the winners as a mixture of fair and corrupted participants. We follow and slightly modify two-stage DEDPUL procedure proposed in [2] for general purposes.

*Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged. Full version is available at https://arxiv.org/abs/1903.00261. Data is retrieved from ftp://ftp.zakupki.gov.ru/.

EC '19, June 24–28, 2019, Phoenix, AZ, USA

ACM ISBN 978-1-4503-6792-9/19/06.
https://doi.org/10.1145/3328526.3329642

In the first stage we build a classifier that distinguishes the winners from the runner-ups by using features associated with patterns (1), (2) and (3). For a given auction winner, the higher the probability of winning is predicted, the more suspicious the auction is. The key idea is to build such classifier that distinguishes corrupted winners and fair runner-ups well, while staying indifferent in regard to fair winners and runner-ups. This can be achieved by maximizing performance of the classifier on the real auctions while minimizing its performance on synthetic placebo data set of fair auctions. We construct such placebo data set by removing the first-ranked bidders (the true winners) from all the auctions. This way, we obtain a new data set where the second places are treated as the winners and the third places – as the runner-ups (Figure 2).

In the second stage we use the classifier's predictions to estimate the prior probability that a random auction is corrupted, and the posterior probability that a specific auction is corrupted – conditional on the probability of winning that the classifier has assigned to its winner. Specifically, we estimate the probability density functions of the classifier's predictions for both samples of winners and runner-ups and apply the Bayes rule to their ratio. The priors and the posteriors are estimated simultaneously: the priors are chosen such that they are equal to the mean posteriors.

A major caveat is that the classifier may still be able to distinguish fair participants with some success. Moreover, same patterns that reflect bid leakage might be involved. For instance, correlation between winning and bidding last in fair auctions might be a consequence of the honest bidders' attempts to resist bid leakage. Ignoring this caveat may lead to positively biased estimates of bid leakage chances. We account for this caveat by using the placebo data set during the second stage. To achieve this, we make an assumption regarding similarity of the fair auctions, real and synthetic:

PARITY: The difference between the probability density functions of the winners and the runner-ups in the real fair auctions is equal to this difference in the placebo auctions.

Parity assumption is a less restrictive analogue to the assumption of independence of bids and timing made by [1]. We find both empirical and theoretical evidence that independence assumption does not hold. In addition, we find empirical evidence that parity holds.

We estimate the prior probability of bid leakage as 16%. We also find that the bid leakage is more likely in auctions with a higher reserve price, lower price fall, where the winning bid is received in the last hour before the deadline (Figure 3); certain regions are more prone to bid leakage (Figure 4).

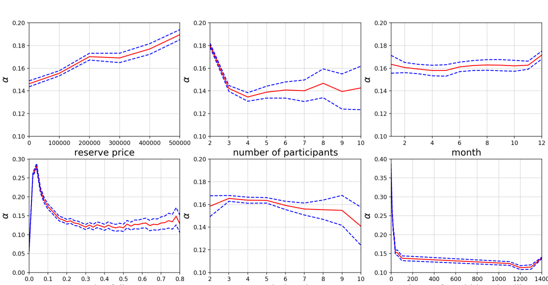

Fig. 3. Bid leakage probability aggregated by auction characteristics

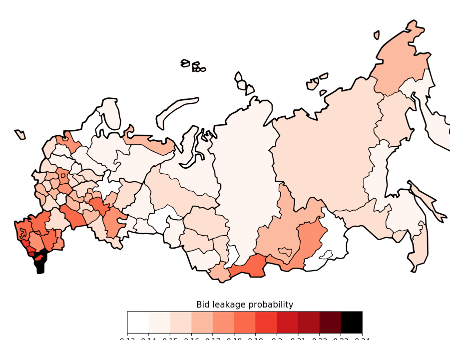

Fig. 4. Regional-wise prevalence of bid leakage

REFERENCES

[1] Pasha Andreyanov, Alec Davidson, and Vasily Korovkin. 2016. *Corruption vs Collusion: Evidence from Russian Procurement Auctions*. Technical Report. mimeo: UCLA.

[2] Dmitry Ivanov. 2019. DEDPUL: Difference-of-Estimated-Densities-based Positive-Unlabeled Learning. *arXiv preprint arXiv:1902.06965* (2019).