Accurate Algorithmic Detection of Collusion With Up to Eight Cartels Using ARIMA and Random Forest Methods

Benjamin Thomas

June 10, 2024

Abstract

Average price auctions are a common technique used for procuring public works contracts since they may help to avoid firm cost cutting or breach of contract due to firms underestimating their own costs and thereby bidding too low. However, a major drawback of average price auctions is that collusive behavior appears indistinguishable from bad play, especially when multiple bidding rings are present in an auction. As a result, current detection methods have to rely on evidence of coordination between auctions, which is often tenuous. I use a dataset of Italian road procurement auctions in Turin where a court case made bidding rings known and use ARIMA to detect how many bidding rings are present. I show that ARIMA can successfully learn the underlying structure of average price auctions, and classified of the residuals succeeds with $p = 0.88$. This suggests that there is fundamental structure in collusion which can be distinguished from regular poor play.

1 Introduction

1.1 Why You Should Care About Average Price Auctions

Public procurement auctions are difficult business. Government agencies have to balance the goals of getting projects done cheaply and getting projects done right. One approach to this is to auction off a contract to build a project to the lowest bidder. Hopefully, with extensive vetting of the bidders and a well specified contract, the resulting project will be completed for cheap, and the resulting road, bridge, tunnel, or whatever it is will be of high quality. Empirical evidence suggests that well specified contracts (Chen et al.) and effective licensing (Decarolis, Moretti and Valbonesi) do protect the quality of the end products. However, there are substantial externalities to these processes. In particular, these processes are slow, expensive, and prone to suffer from corruption, thereby making the market less competitive. Thus, in practice, these measures do not stop the endemic moral hazard problems of first price auctions to a satisfactory degree (Hojin). Further, the price benefits of first-price auctions are often hampered by collusion. For example, see Kawai and Nakabayashi's recent blockbuster paper, which discovered over a third of Japan's construction procurement auctions were fixed. This is not to suggest that first price auctions are a lost cause. Far from it. Padhi et al, for example, proposes numerous ways to design auctions as to deter collusion, and many variants of first-price auctions, such as Vickrey auctions, reduce the risk of moral hazard substantially and, surprisingly, have similar bid distributions to average price auctions (Chang et al.), although, of course, different winners.

An auctioneer may, understandably, be far more concerned with moral hazard than sticker price. This maps squarely to the intuition that, if we want to build a house, we probably do not want to hire the cheapest construction company. Surely they are cheap for a reason and would charge more if their service was a higher quality! As a result, many countries and jurisdictions use average price auctions, among them Italy, Malaysia, Taiwan, and the state of New York. However, despite this, there is a dearth of research on average price auctions. This is due, in large part, to the mistaken assumption of many economists that any auctioneer who is running an average price auction must be making a mistake, since microeconomic theory suggests that average price auctions are always less efficient than first price auctions. This, however, ignores the various real-world consequences of moral hazard. This confusion is perhaps best displayed by a debacle in 2008. After the European Union passed regulation which forced Italy to force to first price procurement auctions, bidding rings quickly took over the market, and when the regulation was repealed in 2011, Italy immediately transitioned back to average price auctions (Conley and Decarolis).

1.2 What We Know About Collusion

The unfortunate result of the lack of interest in average price auctions is that very little work on the detection of collusion in average price auctions has been performed. In fact, the only substantial work on this subject was by Conley and Decarolis, which applied traditional methods collusion detection methods. The main method employed testing whether firms participation in auctions seemed to be coordinated in a non-random way. However, this method is unsurprisingly subject to large amounts of noise. I leverage a convenient fact about average price auctions: collusion is highly visible. Consider the following example of collusive behavior in an average price auction:

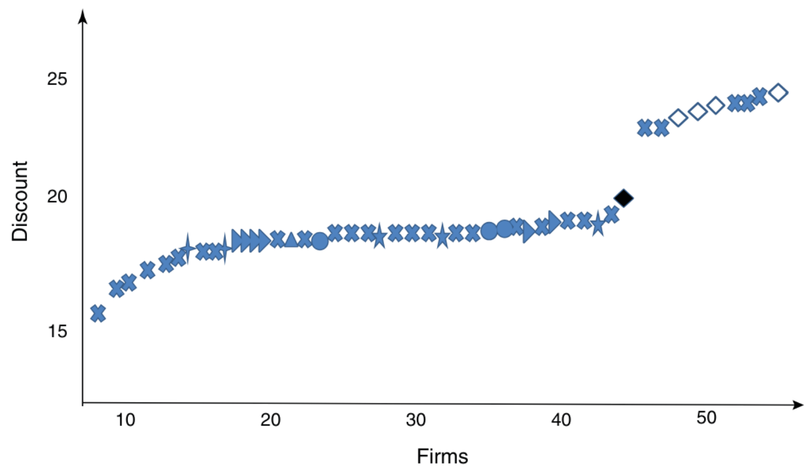

Figure 1: Visualization of "bid steering" in an average price auction from Conley and Decarolis. Shapes correspond to cartel membership; the goal of the bid steering is to make the dark rhombus win.

Since the average bidder wins, and the Nash equilibrium behavior in an average price auction is bid at a publicly known mode, firms participating in bidding rings have a designated winning firm bid either slightly above or below the mode and then have a number of firms "assist" by betting substantially away from the mode. While this behavior is quite visible, it is hard to work with, since multiple bidding rings may try to "steer" the mean either up or down by idiosyncratic amounts. Further, low skill bidders may simply bid poorly. As a result, we cannot simply circle off the mode bids and declare them colluding. Even if they are colluding with *some firm*, they may not be colluding with each other!

The practical consequence of this is that, although it was an open secret for many years that many Italian procurement auctions were rigged, the bidding distributions alone were not sufficient to prosecute. However, police quickly discovered that bidding rings were usually built around obvious family or social ties and successfully pressed charges. While collusion has not stopped after these prosecutions, it does appear to have become more sophisticated (Conley and Decarolis).

2 Research

2.1 Research Question

I test whether it is possible to use time series analysis to determine how many bidding rings are in an average price auction. I test ARIMA models directly and the efficacy of fitting random forest classifiers to ARIMA residuals.

2.2 Data

I use a dataset collected by Conley and Decarolis, which comes from the legal office of the municipality of Turin. The dataset consists of 276 APAs conducted by Turin between 2000-2003 to procure roadwork jobs of homogeneous cost and scale. As discussed in section 1.2, in 2008, Turin convicted many managers of these construction firms for colluding in the APAs. As a result, we know of 8 bidding rings as well as the firms present in each bidding ring. While there is substantial heterogeneity between the number of firms in each cartel, as well as each firm's high-level strategic behavior (for example, deciding when to participate in a particular auction and, if so, how many firms to send), these are analyzed to great effect by Conley and Decarolis. However, given participation in an auction, I find that the underlying data is not very sensitive to these higher-level strategic decisions. This data also includes the discount each firm bid in any given auction.

An important weakness of this dataset is I do not have access to a truth table of which auctions collusion took place in, only which firms were in which cartels. Therefore, I expect there to be much noise in the underlying data resulting from this fact.

2.3 Auction Rules

The rules of the relevant APAs are as follows. First, firms must have appropriate licensing to enter. It is possible to enter with licensing that is only sufficient for some portion of the contract and to sell the rest as subcontracts. Second, all firms anonymously send bids. Third, the average of these bids is taken, μ_1. Fourth, the bottom 10% and top 10% of bids around μ_1 are then eliminated, and μ_2 is calculated from the remaining bids. Fifth, and finally, the lowest bid which is strictly greater than μ_2 is selected.

2.4 Data Preprocessing

I perform three forms of preprocessing on the data. First, I convert auction data into time series. This is done as follows. For all auction j, I obtain the discount bid by each firm i. Then, I create a time series where $t_1 = 1$, and then for each firm i, I add their discounts from smallest to largest to the remaining observations. This has the unfortunate effect of creating auctions of irregular size, precluding seasonality models. The inclusion of the value 1 at t_1 was important, and in fact, the classification failed without it. This is unsurprising, since this is necessary to know whether the smallest discount was near the mode. If not, this would be highly suggestive of collusion.

Next, I desegregate the auction by the suspected number of bidding rings. To successfully desegregate, I check every auction for the number of firms from each known cartel, c. If any cartel has more than n members, I count that cartel as participating. *A priori*, it was unclear what n ought be, since bidding rings could collude with as few as two members, but it seemed unlikely they would attempt to collude in this case. Further, even if two firms colluded, could this be detected? By contrast, if I set n too high, I may exclude important elements of the data. Testing for optimal n is described in detail in 2.5.

Finally, I create "bucket auctions" by appending each auction in the dataset with c suspected cartels to every other auction with c suspected cartels. For example, if auction 1 has 10 bidders and auction 2 has 7 bidders and both have c suspected bidding rings, then I create an auction where t_{12} is given by auction 2's t_1. While this feels unorthodox, the justification is as follows. Just like how adding a 1 to the first observation of each auction was necessary to detect if the lowest discount was associated with the mode, adding a 1 at the end was necessary to assess whether the highest discount was associated with the mode. In theory, a 0 could also do this, but this has substantially worse results. Given that all auctions had to start and end with 1, I originally tried "bucket auctions" as an experiment. However, this substantially increased performance, likely because it allowed the model to learn something about the appropriate distance between the highest and lowest discount.

2.5 Results and Discussion

First, I attempt classification using ARIMA. The method is as follows. For $n \in [1, 5]$, perform data preprocessing. Then, fit an ARIMA model to each bucket auction. I test all ARIMA models that are somewhere between ARIMA(1, 0, 1) and ARIMA(3, 2, 3) and pick the model with the best AIC score. Note that BIC returns the same choice of ARIMA model, and the mode ARIMA configuration was ARIMA(2, 1, 2), although various models were used. Then, for each auction in the data set, we cross validate its values with each ARIMA model and guess that the ARIMA model which had the lowest mean squared error is associated with the correct number of collusive firms.

Before presenting the results, there are a couple concerns I want to address. First, since each model was trained on every positive instance, should we be worried about over fitting? I believe not. Since we select using AIC and limit the complexity of models to (3, X, 3), this would be difficult. Using every positive instance is additionally useful, since this reduces classification issues do to noise in the data. Next, is this sensitize to the order of the auctions in the bucket auction? This is a reasonable concern, since there are many orders we could append the auctions together. However, I find that this has no significant effect. Finally, why mean square error? While I tested against mean absolute error and against mean absolute percentage error, both has substantially worse results. This is likely because the strongest evidence of collusion was in outliers from steering bids, so weighing these higher aided in identification.

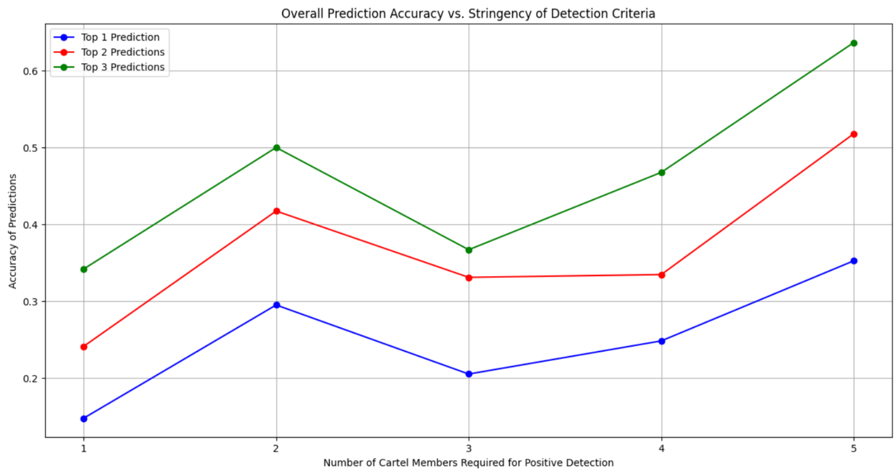

Figure 2: The Classification Accuracy of Naive ARIMA. n is given by the X axis. Results of allowing two or three guesses included for demonstration.

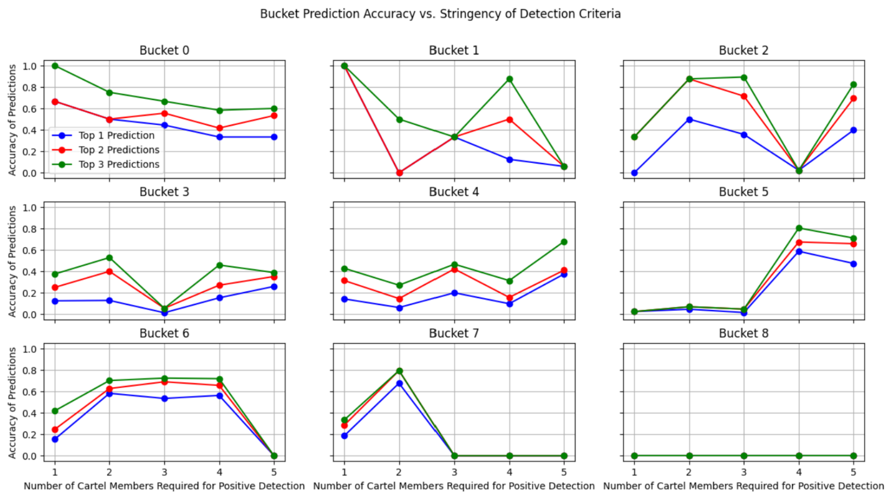

Figure 3: Desegregation of results by number of cartels in an auction (bucket size), and stringency of detection given by n.

There are a few notable features of these results. Most notably, when we set the detection of cartels to $n = 1$, the ARIMA models perform identical to randomly guessing. This suggests the ARIMA are actually learning something about the structure of collusion, rather than a correlated factor, like the number of bidders in an auction, since there is no false positive in a situation where we would not expect to see the effects of collusion, but would expect to see the effects of most correlated factors. This gives us confidence.

It is also important that there are two optima in the figure 2. $n = 2$ and $n = 5$. Note that larger plots yield that $n = 6$ performs worse than $n = 5$. This suggests that $n = 5$ has the optimal mix of not including false positives for bidding rings with true positives for bidding rings, while $n = 2$ is effective since it captures all the relevant data, subject to substantial noise.

Examining figure 3 reveals two important trends. First, as the detection criteria increase, the viable space decreases. Therefore, $n = 2$ may perform better than $n = 5$ in the important sense that the baseline performance of guessing is lower when $n = 2$. We also see phenomenon where certain buckets fail all detection at certain n values. I speculate that this is because, subject to noise, adjacent buckets may appear quite similar, so we would expect this to happen in a number of cases.

Finally, we consider the fitting a random forest classifier to the residuals of each the ARIMA models and testing the accuracy of the resulting classifications. This is a natural approach, because ARIMA is excellent at understanding the trend of the data, while random forest methods are effective at interpreting features of the data. We should expect the residuals of the ARIMA model to highlight when bids are substantially different from the expected distribution and for random forest to learn the interpretation of these deviations (Wang and Shugang). However, this process is not robust to different parameters of the ARIMA models. While quite slow, hyperparameter testing yields that enforcing ARIMA(2, 1, 2) for all models is optimal for all n. Further, I find that results are not substantially different across n. Since random forest has substantially more memory than ARIMA, I train on a random subset of the data and validate against data which the model has not seen in order to avoid over-fitting. Random forest classification of residuals is correct with $p = 0.88$, although there is some variation by bucket size and based on the random selection of training and validation data.

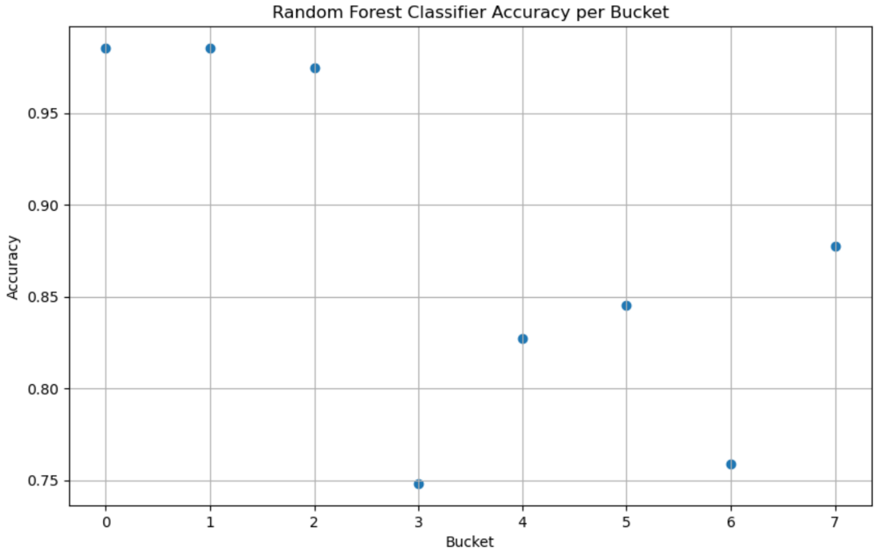

Figure 4: Desegregation of Random Forest classification results (bucket size), $n = 5$.

3 Conclusion

3.1 Parting Thoughts

In this paper I have attempted to motivate why we should be interested in more robust detection of collusion in average price auctions and I have shown that Random Forest Classification of ARIMA residuals is a highly effective approach to determine the number of bidding rings in a particular average price auction. However, my current approach is limited in that it does not determine which firms are in the bidding ring. Further, since the equilibrium behavior in an average price auction is to play the mode, and the mode is determined by society in a way that may be random, this approach may be difficult to replicate on a dataset where the bidding rings are not known.

Nonetheless, I believe that even as a proof of concept, this result is exciting, since it suggests that there is underlying structure in average price auctions which, with appropriate methods, can be used to catch bidding rings.

3.2 References

- Conley, Timothy and Francesco Decarolis. 2016. “Detecting Bidders Groups in Collusive Auctions,” American Economic Journal: Microeconomics, 8(2).
- Chang et al. 2015. "An Investigation of the Average Bid Mechanism for Procurement Auctions," Management Science, 61:6, pp. 1237-1254.

- Chen et al. 2009. "Managing Project Failure Risk Through Contingent Contracts in Procurement Auctions," Decision Analysis, 7(1).
- Decarolis, Francesco. 2014. "Awarding Price, Contract Performance, and Bids Screening: Evidence from Procurement Auctions," American Economic Journal: Applied Economics, 6(1).
- Hojin, Jung. 2016. "Renegotiation on Incomplete Procurement Contracts," Applied Economics, 23.
- Kawai, Kei and Jun Nakabayashi. 2022. “Detecting Large-Scale Collusion in Procurement Auctions,” Journal of Political Economy, 130(5).
- Moretti, Luigi, and Paola Valbonesi. 2015. "Firms’ Qualifications and Subcontracting in Public Procurement: An Empirical Investigation," Journal of Law Economics and Organization, 31(3).
- Padhi et al. 2016. "Design of Auction Parameters to Reduce the Effect of Collusion," Decision Sciences, 47(6).
- Wang, Jinghui, and Shugang Tang. “Time series classification based on arima and adaboost”, MATEX Web of Conferences 309, 03024.

References