Predicting the Award Price of First Price Sealed Bid Procurement Auctions

Fabian Blasch

09/26/2022

Motivation: The Importance of Public Auctions

Auctions are a vital tool for governments to procure contracts. For construction contracts, first price sealed bid auctions are of particlar importance.

▶ The authorities of the European Union for example spent around 14% of their GDP on public procurement in 2017 (Rodríguez et al. 2020).

▶ Similar observatios can be made for the U.S. economy, one state of particulat importance for this thesis is Colorado.

▶ In 2021 the Budget for Transportation in Colorado amounted to roughly $2 billion. Out of this Budget the CDOT awarded $790 millions worth of contracts to construct design and repair bridges and highways. All of those contracts were procured via first price sealed bid auctions.

Thesis Overview

▶ Provide an award price prediction model for the Colorado Department of Transportation.
  ▶ This model would enable the auctioning entity to plan their budget more accurately.
▶ Unsupervised Collusion Detection
  ▶ Examine whether the interaction of certain bidders has an effect on award prices. A signigicant interaction effect could allude to bid rigging schemes.

Data: Source

An example of a bid tab, as published on the website of the CDOT.

Letting No:	20151112	Contract ID:	C19868	Project(s): STU1211-084
Letting Date:	November 12, 2015	Region:	1	
Letting Time:	10:00 AM	Contract Time:	260 WORKING DAYS	Counties: JEFFERSON, REGION 1

Contract Description:

SH121(WADSWORTH)-HIGHLAND DR-10TH AVE-JEFFERSON CO

THIS PROJECT IS LOCATED ON WADSWORTH BETWEEN HIGHLAND AND 10TH.

CONSTRUCTION WILL INCLUDE A FULL CONSTRUCTION WITH WIDENING OF ONE
LANE IN BOTH DIRECTIONS, AND A MULTI MODAL TRAIL ON BOTH SIDES. THE
MAINLINE PAVING WILL BE CONCRETE. THE WORK ALSO INCLUDES A CONCRETE
BOX CULVERT NEAR HIGHLAND TO CARRY LAKEWOOD GULCH UNDER WADSWORTH.

CDOT WILL ONLY BE ACCEPTING ELECTRONIC BIDS FOR THIS PROJECT. PLEASE
CONTACT BID EXPRESS CUSTOMER SERVICE AT 1-888-352-2439 TO OBTAIN AN
ACCOUNT IF NECESSARY.

Rank	Vendor ID	Vendor Name	Total Bid	Percent Of Low Bid	Percent Of Estimate
0	-EST-	Engineer's Estimate	$9,821,027.20	91.58%	100.00%
1	870A	SEMA CONSTRUCTION, INC.	$10,723,550.00	100.00%	109.19%
2	884A	HAMON INFRASTRUCTURE, INC.	$10,817,000.00	100.87%	110.14%
3	1275A	CASTLE ROCK CONSTRUCTION COMPANY OF COLORADO, LLC	$10,817,845.03	100.88%	110.15%
4	065A	CONCRETE WORKS OF COLORADO INCORPORATED	$11,614,565.78	108.31%	118.26%
5	232A	AMERICAN CIVIL CONSTRUCTORS, INC. dba ACC Mountain West	$12,338,888.00	115.06%	125.64%

Figure 1: Bid Tab Example

Data: Extraction

- The following text based data was extracted utilizing the package *pdftools* and regular expressions (Ooms 2022):
    - Contract ID
    - County
    - Contract Time
    - Contract Description
- The table containing the bids, the unique bidder intentifiers and the engineer's estimate was extracted utilizing the R package *tabulizer* (Leeper 2018).

Data: Final Dataset

The final dataset features 430 observations and 1086 variables.

- ▶ County
- ▶ Letting month
- ▶ Letting year
- ▶ Contract time
- ▶ Number of bidders
- ▶ Engineer's estimate
- ▶ Award price
- ▶ 169 binary variables, representing the bidder identities
- ▶ 652 binary variables, representing pair-wise bidder interaction terms
- ▶ 258 binary variables, representing the contract description hit words

Methods

For the Prediction Model the following models were applied utilizing different preprocessing schedules:

- ▶ Elastic net
- ▶ Random forest
- ▶ eXtreme Gradient Boosting
- ▶ OLS estimated linear model

Methods: Elastic Net

An elastic net is a linear model that utilizes ℓ_1 and ℓ_2 regularization, we may forumalte the elastic net as a constrained minimization problem,

$$\hat{\beta} = \arg \min_{\beta} |\mathbf{y} - \mathbf{X}\beta|^2,$$

subject to,

$$(1 - \alpha)|\beta|_1 + \alpha|\beta|^2 \le t.$$

Where $\alpha \in [0, 1]$ and $t \in \mathbb{R}^+$.
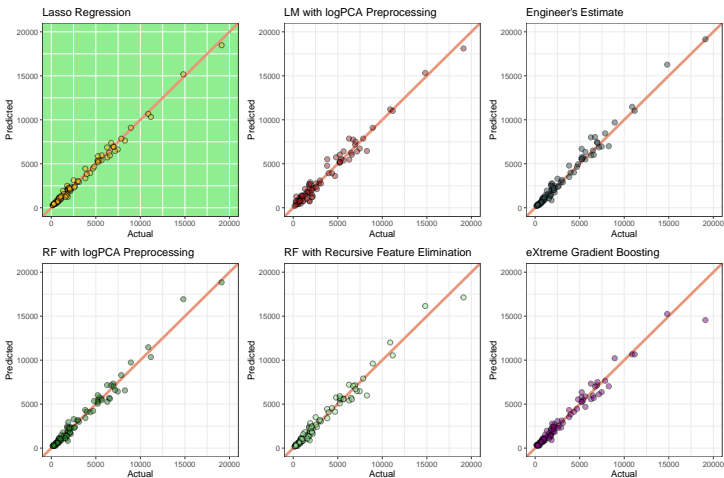
Results: Out of Sample Performance

Figure 2: Performance Comparison

Results: Out of Sample Performance

The following table lists the performance of the applied methods utilizing linear and quadratic loss functions, i.e.,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y} - y)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y} - y|.$$

	Lasso	Eng. Est.	logPCA_RF	rfe_RF	XGB	logPCA_LM
RMSE	326.1261	497.0567	509.7934	560.7600	671.6634	609.4673
MAE	241.7894	327.9388	348.2869	373.1132	376.3191	448.2662

Unsupervised Colusion Detection: Post-Selection Inference

References

Leeper, Thomas J. 2018. *Tabulizer: Bindings for Tabula PDF Table Extractor Library*.

Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*. https://CRAN.R-project.org/package=pdftools.

Rodríguez, Manuel J. García, Vicente Rodríguez Montequín, Francisco Ortega Fernández, and Joaquín M. Villanueva Balsera. 2020. "Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain." *Complexity* 2020: 1–20. https://doi.org/10.1155/2020/8858258.