over all good job. there is room to improve the narrative. I have made some suggestions below.

9/10

Econ 4567 Auction Theory: Caltrans Paper

Genevieve Mendoza and Claire Gottreich

February 21, 2023

This paper analyzes data from 705 procurement auctions held by the California Department of Transportation (commonly known as CalTrans).

1 Data

1.1 Institutional Details

The California Department of Transportation (Caltrans) is the department that manages Aeronautics, Highway Transportation, Mass Transportation, Transportation Planning, Administration, and the Equipment Service Center in California. To outsource the labor for their highway construction projects, Caltrans runs low-bid procurement auctions. Within the auctions, there are typically large and small business bidders. Because small businesses have lower economies of scale, their costs are higher compared to large businesses.

Due to the difference in costs, Caltrans grants "bid preferences" to the small business bidders. Small businesses must meet three qualifications to be obtain the Small Business Certification. The certification requires that the business must be independently owned and operated in California, have no more than 100 employees, and over the last three tax years can only earn under $10 million average annual gross receipts. The benefit of the Small Business Certification is that there is a higher probability of winning the auction.

Once all bids are submitted, the lowest bid wins the auction. However, if a small business's bid is within 5% of the lowest bid, they win the auction and are awarded the contract for construction. While the 5% discount is used to determine the winner, it is not applied to the actual amount the business is paid for the project: Caltrans will pay the true price the winner bid.

1.2 Data Overview

what do you mean? and this is too informal

One immediate takeaway from the summary statistics is that small businesses are often refusing to bid on larger contracts, given that their bids are both much smaller on average and have a smaller standard deviation.

where do I see this? if you claim something, you have to show it

In addition, out of 705 auctions in total, there were 5.7 bidders in each auction on average, and the average bid was $968,241.65.

what about auctions with 1 bidders?

We calculated the winning bids and saw that on average, the winning bid was $39,417 lower than the state's cost estimate. This is a relatively small difference, but it does suggest that competition among the bidders does drive the procurement cost down for

?

drop auctions with n=1

Add Caption to Tables (and Figures)

	Mean	Standard Deviation	Minimum	Maximum
Bids	986241.65	3311628.21	44655.00	58547700.00
Small Business Bids	531334.61	723168.10	49650.00	15485561.50
Number of Bidders	5.70	3.18	1.00	20.00
? Business types present	1.64	0.48	1.00	2.00
Engineer's Estimates	943980.19	3734611.68	74000.00	60058000.00
Workdays	94.28	155.89	8.00	1430.00

whenever the numbers are large, use millions

Add Note.

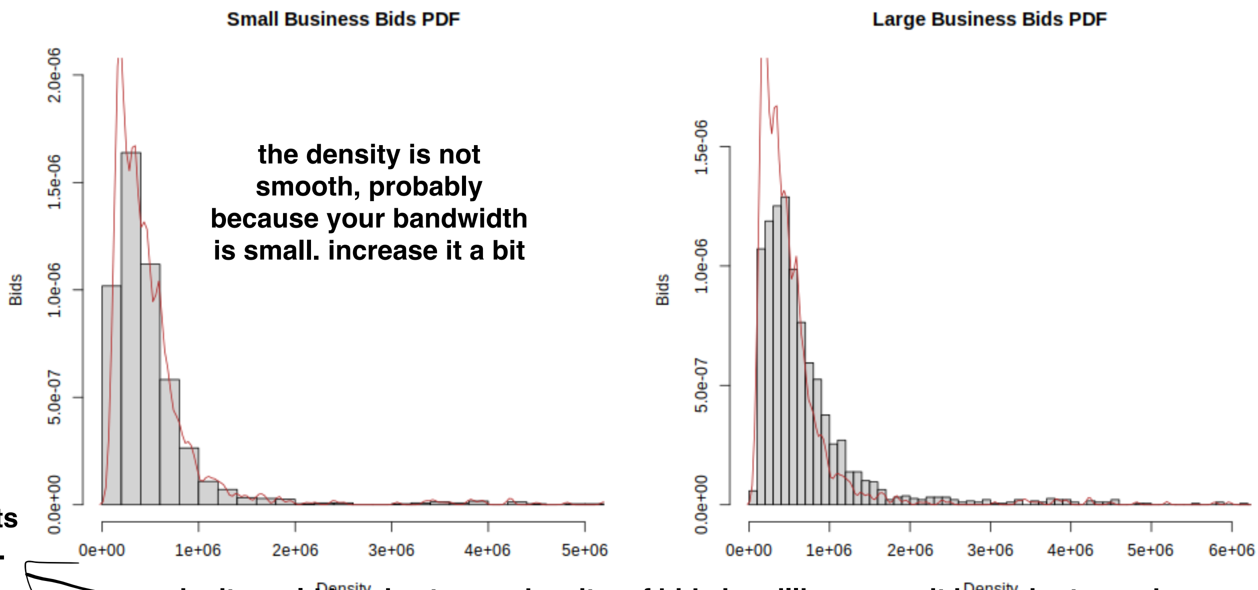
the state. It could also potentially represent a winner's curse if the cost estimates are very accurate, but this would require more analysis to confirm. Furthermore, the standard deviation of the differences is quite high at almost $1 million, and there are some cases where Caltrans is forced to pay a premium over their estimate - it is unclear if this represents unreliability by the engineers making the estimates, or an underlying dynamic of the auction process.

another interesting variable would be to present bids/estimate.

1.3 Bidding Behavior

1.3.1 Kernel Density Estimation

First we estimated the true distribution of bids for large and small businesses using kernel density estimation with a bandwidth selected via leave-one-out cross-validation. Kernel density estimation relies on the second derivative of the density function, and will as a result tend to overestimate peaks and underestimate valleys. We can see that in this plot, where the density estimate is much higher than the first peak.



the density is not smooth, probably because your bandwidth is small. increase it a bit

the spikes could be artifacts of bandwidths.. thats why the PDFs of log bids in next page are smoother

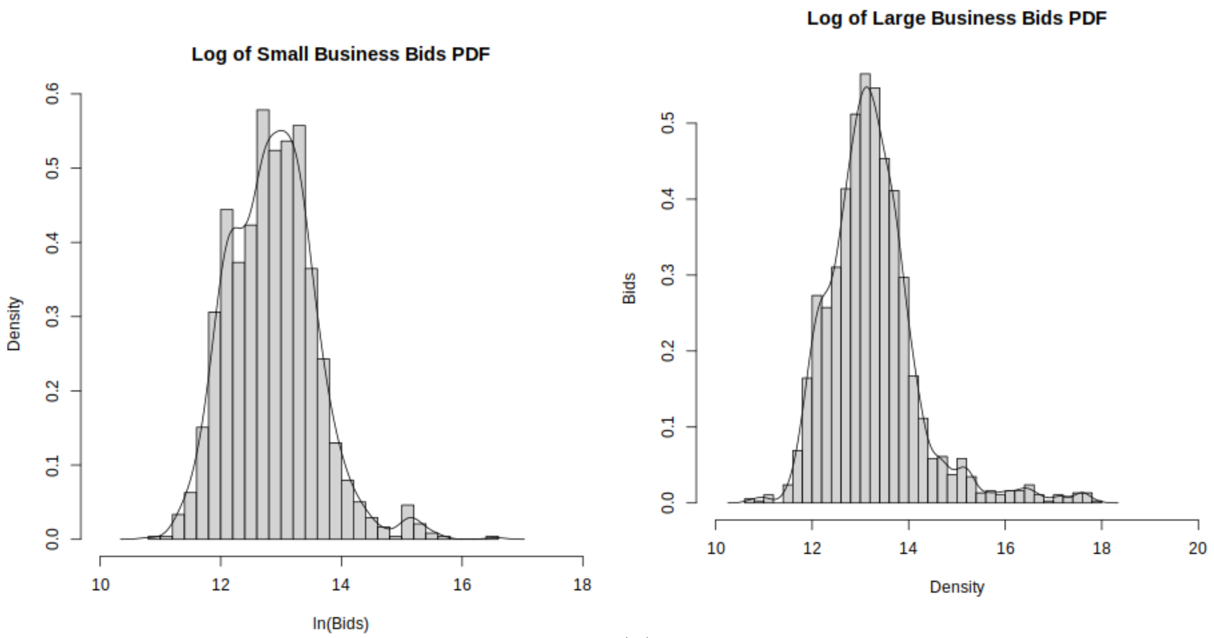
again, it would be nice to see density of bids in millions... so it is easier to read.

The large business bids do have a longer tail, which reflects that they are more able to take on projects with a large cost than the smaller businesses. There are also clusters in the bid frequencies, which we can see in the spikes in the KDEs of the PDFs. It makes sense that the distribution of bids is not perfectly uniform, but instead clusters because projects can be grouped into similar size categories, and because people generally prefer to deal with round numbers when it comes to money. Many more projects will receive

not sure what this means

bids at \$1 million than at \$1,010,000.

Because of the data's skew, we applied a log transformation to normalize the data. Log scaling data can effectively compress very wide ranges down, so that extreme bids are proportionally less large. Consequently, the goal of this transformation was to provide a better visualization of the long tail in our data. This reveals a very slight bimodal peak

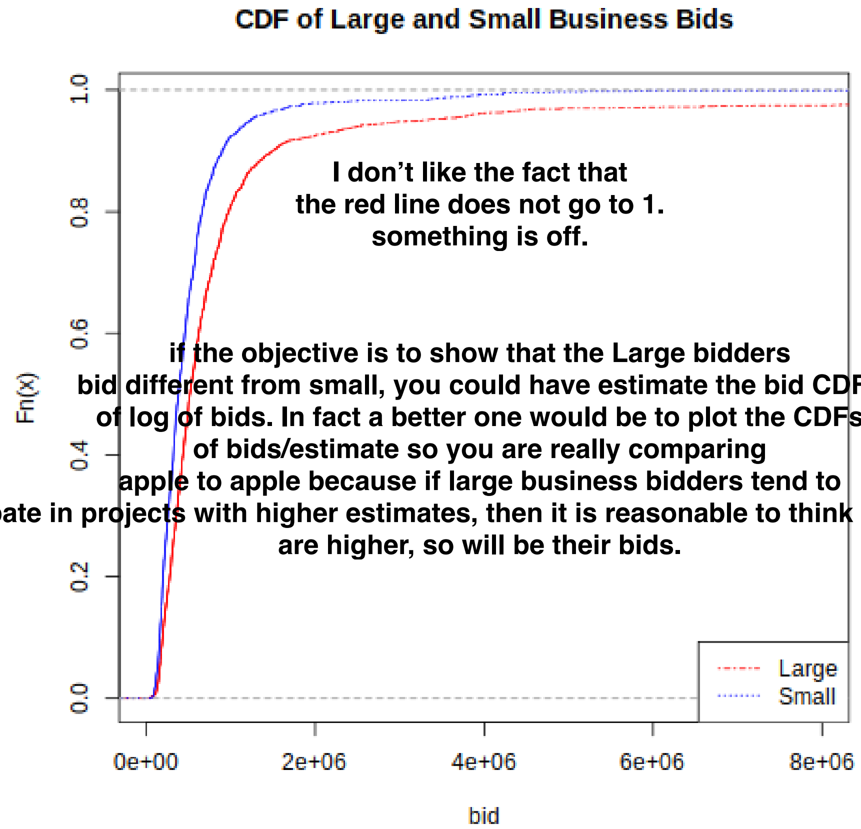


(a) PDF of the log-bids by small businesses

(b) PDF of the log-bids excluding small businesses

around bids of about \$3 million (exp(15)), which might be a common project cost. The larger businesses look to have a very similar distribution, but just have a higher mean, and their tail also extends further up to bids of around \$7-8 million.

We also plotted the CDFs to see patterns in the bid sizes between the two types of bidders. The smaller average bids of the small businesses are very clear in this plot, as the line is strictly greater than the large businesses' estimated CDF.



I don't like the fact that the red line does not go to 1. something is off.

if the objective is to show that the Large bidders bid different from small, you could have estimate the bid CDFs of log of bids. In fact a better one would be to plot the CDFs of bids/estimate so you are really comparing apple to apple because if large business bidders tend to participate in projects with higher estimates, then it is reasonable to think that their costs are higher, so will be their bids.

1.3.2 Regressions

To better understand the bidding behavior of business, we ran regressions on bidders, engineer's estimate, and work days. Bidders is how many bidders there were for a certain project, engineer's estimate is the estimate on how much a professional believes the project costs, and work days is how long the project will take. The first regression we ran included both small and large businesses, and below is the result. **Show me the regression models not just the table.**

use \begin{center} to center the table. and end it with \end{center}

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	224239.4332	32012.8642	7.00	0.0000
num_bidders	-18815.3332	4617.4432	-4.07	0.0000
Estimate	0.8494	0.0041	205.05	0.0000
WorkDays	715.9631	99.6008	7.19	0.0000

add Caption and Notes

show the adjusted R^2
number of observations

The above table shows the output of three regressions of the number of bidders, estimated cost, and workdays spent on the job, on the procurement bids. The first covers all of the data, the second only certified small businesses, and the third large businesses. The results of this regression are as we expected. The negative sign on bidders is consistent with the theory ~~we studied~~, because with more bidders in a particular auction, it drives down the price of the bid. Similarly, as the engineer's estimate increases, the bid should increase because there is a higher cost for the project; and as work days increase, so should the bid because projects that take longer become more expensive.

The results for small businesses also have the same signs as we expected, for the same reasons as above. The only difference in this regression is the degree to which work days

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	224239.4332	32012.8642	7.00	0.0000
num_bidders	-18815.3332	4617.4432	-4.07	0.0000
Estimate	0.8494	0.0041	205.05	0.0000
WorkDays	715.9631	99.6008	7.19	0.0000

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	224239.4332	32012.8642	7.00	0.0000
num_bidders	-18815.3332	4617.4432	-4.07	0.0000
Estimate	0.8494	0.0041	205.05	0.0000
WorkDays	715.9631	99.6008	7.19	0.0000

influences the bid. The coefficient for the regression with only small business bidders is 154.3 compared with 716 for the regression including all businesses. This might indicate that number of work days does not increase the bid as much because smaller businesses might not be able to handle the cost of larger projects that require more time and capital. Therefore, they can't bid as high.

A notable difference in the final regression, which is limited to businesses that tend to be larger, is the greater coefficient for work days, such that there is a much larger coefficient on work days for large business bidders. This can be explained by the fact that large businesses can handle the higher costs of lengthier projects. Their economies of scale allow for them to afford projects that require more time and capital. Because of this, they are more likely to pursue such projects and can bid higher for them.

2 Appendix - Code

not needed right now.

The following is the R code used to prepare the data section of the paper.

```r
load("./src/code/Caltrans_Data/caltransdata.RData")

# summary stats on the data from 705 auctions
# here, getting the number of types of bidders (either 1 or 2)
types_bidders <- vector(length =
   length(unique(caltransdata$ProjectID)))
auction_rows <- match(unique(caltransdata$ProjectID),
   caltransdata$ProjectID)
auction_indexer <- 1
for (auction in auction_rows) {
  types_bidders[auction_indexer] <-
    (caltransdata$NumberofSmallBusinessBidders[auction] != 0) +
    (caltransdata$NumberofLargeBusinessBidders[auction] != 0)
  auction_indexer <- auction_indexer + 1
}

sb_bids <- caltransdata[caltransdata$SmallBusinessPreference == 1,
   ]$Bid
num_bidders <- caltransdata$NumberofSmallBusinessBidders +
```

```r
  caltransdata$NumberofLargeBusinessBidders

my_smry <- function(x) {
  return(c(mean(x), sd(x), min(x), max(x)))
}
all_bids_smry <- my_smry(caltransdata$Bid)
sb_bids_smry <- my_smry(sb_bids)
num_bidders_smry <- my_smry(num_bidders)
num_types_bidders_smry <- my_smry(types_bidders)
eng_est_smry <- my_smry(caltransdata$Estimate)
workdays_smry <- my_smry(caltransdata$WorkDays)
summary_stats <- matrix(Reduce(c, list(all_bids_smry, sb_bids_smry,
                                       num_bidders_smry,
                                       num_types_bidders_smry,
                                       eng_est_smry,
                                          workdays_smry)),
                        nrow = 6, byrow = TRUE)
colnames(summary_stats) <- c("Mean", "Standard Deviation",
   "Minimum", "Maximum")
rownames(summary_stats) <- c("Bids", "Small Business Bids", "Number
   of Bidders",
                             "Business types present",
                             "Engineer's Estimates", "Workdays")
library(xtable)
print(xtable(summary_stats), latex.environments = NULL, booktabs =
   TRUE,
      file = "./src/sections/data-summary.tex")

# how close are the (winning) bids to the engineer's estimate?
find_winning_bid <- function(x) {
  low_bidder <- which.min(x$Bid)
  low_bid <- min(x$Bid)
  if (x$SmallBusinessPreference[low_bidder] == 1 ||
      sum(x$NumberofSmallBusinessBidders) == 0) {
    return(low_bid)
  }
  sb <- x[x$SmallBusinessPreference == 1, ]$Bid
  low_sb_bid <- min(sb)
  if (low_sb_bid / low_bid < 1.05) {
    return(low_sb_bid)
  } else {
    return(low_bid)
  }
}
winning_bids <- by(caltransdata, factor(caltransdata$ProjectID),
   find_winning_bid)
mean(winning_bids - caltransdata[auction_rows, ]$Estimate)

# pdfs and cdfs

pdf_sb <- density(sb_bids, bw = "UCV")
```

```r
png("src/imgs/sb-pdf.png")
hist(sb_bids, breaks = 100, freq = FALSE, main = "Small Business
   Bids PDF",
     xlab = "Density", ylab = "Bids", xlim = c(0, 5e6), ylim = c(0,
        2e-6))
lines(pdf_sb, col = "firebrick")
dev.off()

lb_bids <- caltransdata[caltransdata$SmallBusinessPreference == 0,
   ]$Bid
lb_bw <- bw.ucv(lb_bids, lower = 1e-6, upper = 1e6)
pdf_lb <- density(lb_bids, bw = lb_bw)
png("./src/imgs/lb-pdf.png")
hist(lb_bids, breaks = 800, freq = FALSE, main = "Large Business
   Bids PDF",
     xlab = "Density", ylab = "Bids", xlim = c(0, 6e6), ylim = c(0,
        1.8e-6))
lines(pdf_sb, col = "firebrick")
dev.off()

log_lb_bids <- log(lb_bids)
png("./src/imgs/log-lb-pdf.png")
hist(log_lb_bids, breaks = 30, freq = FALSE,
     main = "Log of Large Business Bids PDF", xlab = "Density",
        ylab = "Bids",
     xlim = c(10, 20))
pdf_log_lb <- density(log_lb_bids)
lines(pdf_log_lb)
dev.off()

log_sb_bids <- log(sb_bids)
png("./src/imgs/log-sb-pdf.png")
hist(log_sb_bids, breaks = 30, freq = FALSE,
     main = "Log of Small Business Bids PDF", xlab = "ln(Bids)",
     ylab = "Density", xlim = c(10, 18))
pdf_log_sb <- density(log_sb_bids)
lines(pdf_log_sb)
dev.off()

cdf_lb_bids <- cumsum(lb_bids) / sum(lb_bids)
cdf_sb_bids <- cumsum(sb_bids) / sum(sb_bids)

png("./src/imgs/cdf.png")
# CDFs of the bids for small and large businesses
plot(ecdf(lb_bids), xlim = c(0, 8e6), col = "red", lty = 4,
     main = "CDF of Large and Small Business Bids", xlab = "bid")
lines(ecdf(sb_bids), col = "blue", lty = 3)
legend("bottomright", legend = c("Large", "Small"),
       col = c("red", "blue"), lty = c(4, 3))
dev.off()
```

```
# regressions
calt_full <- cbind(num_bidders, caltransdata)
full_reg <- lm(Bid ~ num_bidders + Estimate + WorkDays, data =
   calt_full)
print(xtable(summary(full_reg)), latex.environments = NULL,
   booktabs = TRUE,
      file = "./src/sections/data-regressions.tex")

sb_reg <- lm(Bid ~ NumberofSmallBusinessBidders + Estimate +
   WorkDays,
            data =
               caltransdata[caltransdata$SmallBusinessPreference
               == 1, ])
print(xtable(summary(full_reg)), latex.environments = NULL,
   booktabs = TRUE,
      file = "./src/sections/data-regressions.tex", append = TRUE)


lb_reg <- lm(Bid ~ NumberofLargeBusinessBidders + Estimate +
   WorkDays,
            data =
               caltransdata[caltransdata$SmallBusinessPreference
               == 0, ])
print(xtable(summary(full_reg)), latex.environments = NULL,
   booktabs = TRUE,
      file = "./src/sections/data-regressions.tex", append = TRUE)
```