Estimation of Cost Functions for Participants in Caltrans Infrastructure Auctions and Implications for Cost-Minimizing Auction Structure

Genevieve Mendoza and Claire Gottreich*

May 7, 2023

1 Introduction

In this paper we explore the bidding behavior of businesses in the California Department of Transportation (commonly known as Caltrans) auctions. Caltrans outsources their labor for highway construction auctions by using low-bid procurement auctions. In a low-bid procurement auction, the bidder with the lowest bid wins the auction and the price equals the lowest bid. Within Caltrans' auctions there are two types of bidders: small business bidders and large business bidders. Large businesses have more resources and capital, and thus have a greater advantage in winning auctions. Therefore, to account for this, Caltrans uses bidder preference, such that small businesses will win a particular auction if they are within 5% of a large business bidder.

We begin with a close exploration of the data, in which we further explain how Caltrans operates and analyze the bidding data. This is done through analysis of summary statistics, plots of bidding behavior and a regression analysis of additional variables. We have data from 705 procurement auctions held by Caltrans. This paper is broken up into five sections, which will analyze the auction data; find the theoretically optimal bidding functions according to the rules of Caltrans auctions; recover an equation for estimated business costs according to that bidding function; find estimated densities of the cost distributions according to our empirical data; and use these estimated densities to explore if Caltrans could lower costs by using second-price auctions (SPAs) with reserve prices.

This counterfactual was chosen because of two results. The first, the revenue equivalence theorem, states that the expected costs/revenue are the same for both first and second price auctions as long as they have the same method of assigning outcomes and the same expected payment for participants. While the 5% margins for small businesses cannot be implemented in exactly the same way in second-price auction, we will elaborate a system for per-group reserve prices which should preserve the same outcomes.

Second, although reserve prices are never guaranteed to cut costs in a procurement auction, it is possible as long as the number of auction participants does not decrease. Given this, it is worth studying whether Caltrans and other publicly funded utilities which rely on private contractors could reduce their expenditures by changing their auction choices.

*Paper written for Economics 4567: Auction Theory at Washington University, with feedback and advice given by Professor Gaurab Aryal.

2 Data

2.1 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.

2.2 Data Overview

	Mean	Standard Deviation	Minimum	Maximum
Bids	9.876×10^5	3.33×10^6	4.5×10^4	5.9×10^7
Small Business Bids	5.283×10^5	7.211×10^5	5×10^4	1.5×10^7
Large Business Bids	1.278×10^6	4.19×10^6	4.5×10^4	5.9×10^7
Number of Bidders	5.759	3.155	2	20
Business types present	1.673	0.4696	1	2
Engineer's Estimates	9.475×10^5	3.756×10^6	9.1×10^4	6×10^7
Workdays	94.06	155.9	8	1.4×10^3

Table 1: Summary statistics for subset of auctions with 2 or more bidders.

For the summary statistics, we excluded auctions with only one bidder. All statistics are from that subset of the data. The mean small business bid was \$528,318, and the mean large business bid was \$1,277,866. A possible explanation for this discrepancy is that small businesses were not bidding on the largest contracts, perhaps because they do not have the capacity for multi-million dollar projects. The smaller standard deviation of small business bids also supports this theory.

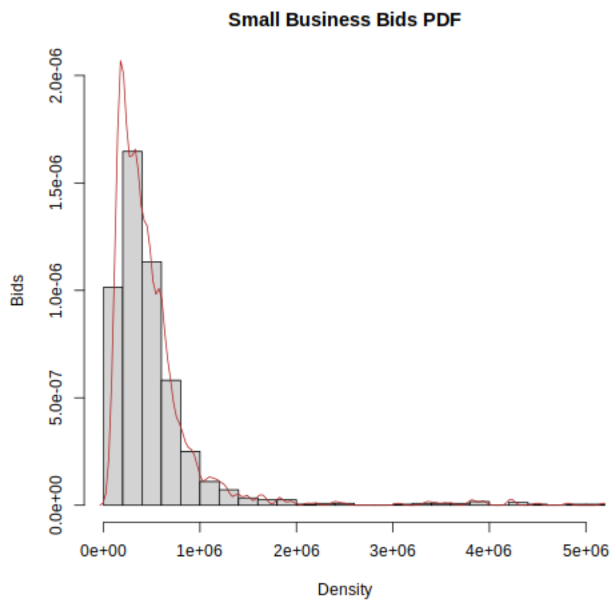
In addition, out of 669 auctions in total, there were 5.8 bidders in each auction on average, and the average bid was \$987,619. The "business types present" statistic refers to whether small business, large business, or both types of bidders participated in each auction. The mean of 1.64 suggests that the majority of auctions had both types participating.

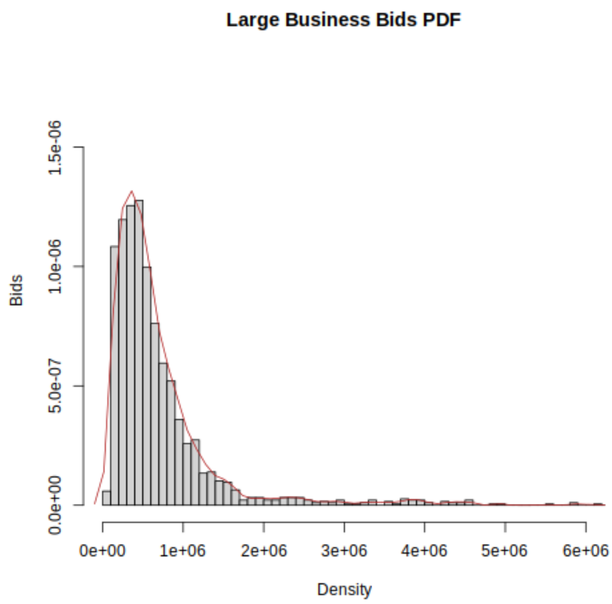
We calculated the winning bids and saw that on average, the winning bid was $52,629 lower than the state's cost estimate. CalTrans was therefore able to reduce costs compared to their own engineers' estimates by running auctions for procurers. Auction theory would suggest that the greater the number of bidders, the more the cost will be driven down for the procurer due to competition. There could potentially be a winner's curse if it is impossible to drive costs much lower than the estimate, but we cannot determine that from this data alone. 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.

2.3 Bidding Behavior

2.3.1 Kernel Density Estimation

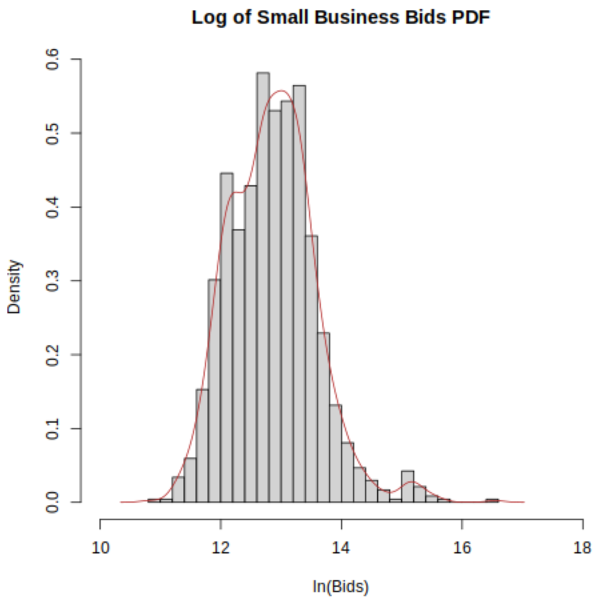
First we estimated the true distribution of bids for large and small businesses using kernel density estimation with a bandwidth selected via leave-one-out cross-validation. Kernel density estimation relies on the second derivative of the density function, and will as a result tend to overestimate peaks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



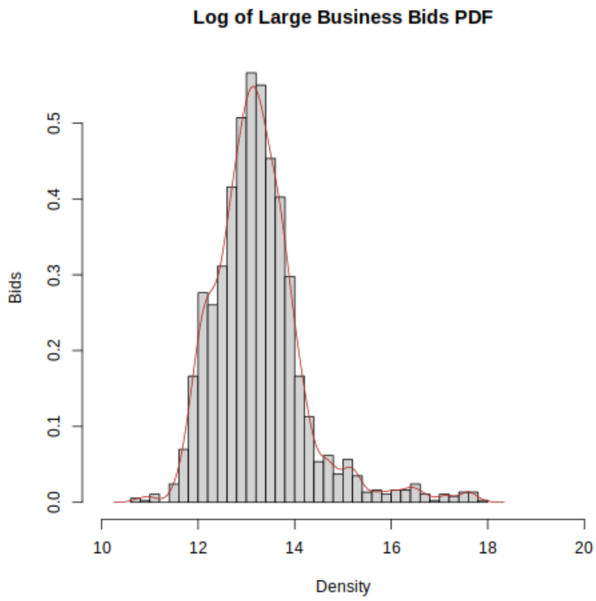


The large business bids do have a longer tail, which reflects that they are more able to take on projects with a large cost than the smaller businesses. There are also clusters in the bid frequencies, which we can see in the spikes in the KDEs of the PDFs. One possible explanation is that the distribution 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 bids at $1 million than at $1,010,000. It is also possible that clusters are simply an artifact of the KDE process, but we used biased cross-validation to select the bandwidths, which should not undersmooth the data.

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



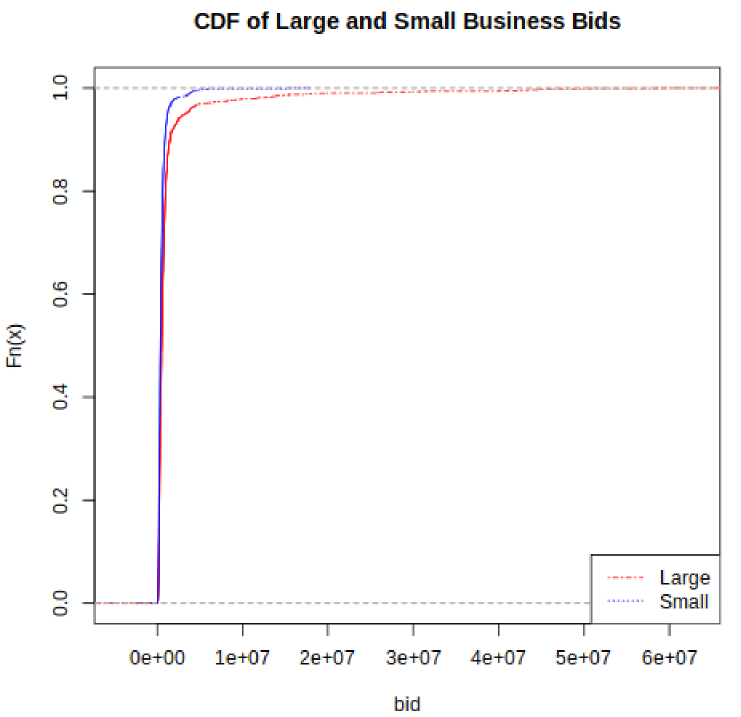
(a) PDF of the log-bids by small businesses

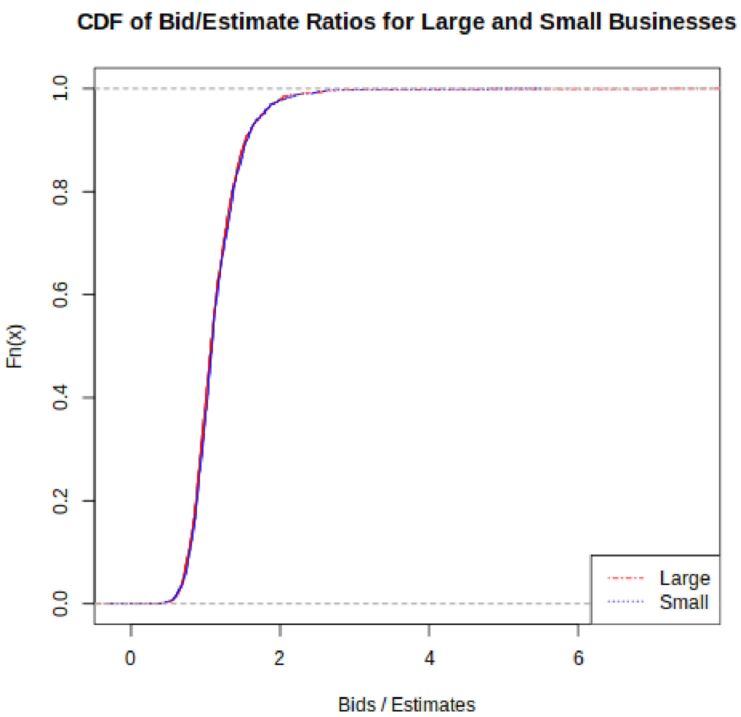


(b) PDF of the log-bids excluding small businesses

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. Lastly, we plotted the CDF of the bid-estimate ratio for only small and large businesses. Although the above plot showed the distribution of large business bids had a much longer tail than small businesses, they generally appear to bid a similar ratio of the estimate. This also suggests that large businesses are taking on bigger projects, because when costs are higher (where estimate is related to the business's true cost), they bid higher just as small businesses do. It also is interesting that because of the 5% rule, small businesses could theoretically 'overbid,' but they are also bidding at the same ratio of the estimate. This is likely because they have less economies of scale and want to remain as competitive as possible, even with the 5% rule.





2.3.2 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. We ran regressions on all businesses, small businesses only, and large businesses only. The first regression including both small and large businesses gave the equation:

$$\text{Bids} = 222576 - 18609\#\ \text{Bidders} + 0.8493\text{Estimate} + 718\text{Workdays}$$

The regressions on a subset gave:

$$\text{Bids} = 136254 - 16423\#\ \text{Bidders} + 0.9469\text{Estimate} + 149\text{Workdays}$$
$$\text{Bids} = 178043 - 20918\#\ \text{Bidders} + 0.8396\text{Estimate} + 1370\text{Workdays}$$

In addition, below we have included the regression tables for further analysis.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	222576.4332	32736.1113	6.80	0.0000
num_bidders	-18609.7317	4701.4001	-3.96	0.0001
Estimate	0.8493	0.0042	203.89	0.0000
WorkDays	718.0144	100.7270	7.13	0.0000

Table 2: Regression on Bids, all data

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	136254.2725	13161.1324	10.35	0.0000
NumberofSmallBusinessBidders	-16423.0841	2467.3768	-6.66	0.0000
Estimate	0.9469	0.0090	104.76	0.0000
WorkDays	149.0827	37.4165	3.98	0.0001

Table 3: Regression of Small Business Bids Only

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	178043.7249	59430.0482	3.00	0.0028
NumberofLargeBusinessBidders	-20918.5826	14608.5652	-1.43	0.1523
Estimate	0.8396	0.0057	147.16	0.0000
WorkDays	1370.7955	182.1543	7.53	0.0000

Table 4: Regression on Large Business Bids Only

The above tables show the output of three regressions of the number of bidders, estimated cost, and workdays spent on the job, on the procurement bids. The results of this regression are as we expected. The negative sign on bidders is consistent with the theory, 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 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.

3 Model

3.1 Assumptions

First, recall the auction rules: a small business bid within 5% of the lowest bid will win, and the procurer will receive their bid. Thus, the 5% rule affects the expectation of winning, but not the payoff. Because there are two groups of businesses, we have asymmetric bidders which we will refer to as L and S, with $n_L + n_S = n$, and $n \geq 2$.

Similarly, refer to the CDFs of the costs as $F_L \neq F_S$ on the support $[\underline{c}, \bar{c}]$.

We assume that during the CalTrans auction, the bidders had costs that were independent and drawn from an identical distribution, and that the costs were private. (Since these were procurement auctions, cost is the analogue to valuation). However, the distribution is conditional on some estimate.

These distributions are unknown, and we do not have complete information, on their parameters. However, we do know that they are conditional on an engineer's estimate provided by CalTrans, which approximates the true cost. distributed according to some unknown, distinct parameters, Therefore, we write that $c \sim F_{L \text{ or } S}(\cdot \mid E)$, where $\cdot$ represents parameters and E represents the estimate.

Bidding functions are assumed to be monotonically increasing in cost. Similarly, all bidding functions are assumed to satisfy $\beta(0) = 0$. Require that $1.05\beta_L(\bar{c}) = \beta_S(\bar{c})$, and that each β_i has an inverse $\phi_i = \beta_i^{-1}$.

Finally, we assume that within the groups, each bidder uses the same optimal strategy, denoted β_S and β_L.

3.2 Bidders' Optimization Problems

Since the auction is a first price auction, we can define the expected payoffs for a small business i. Π refers to the profit or expected payoff for i. The numbers of small/large bidders participating in the auction are n_S, n_L). Index the large business competitors with k and other small business competitors with $j \neq i$. Recall that for each opponent, $b_j = \beta_S(c_j)$ and $b_k = \beta_L(c_k)$. Since valuations are private, we simply refer to opponent bids b_j, b_k in the below equation. The probability on the right side is that all $n_S - 1$ SB costs are higher than their cost and that all n_L LB costs are higher than $\frac{1}{1.05}$ their cost.

Since costs are private, we estimate this with the inverse bid functions.

$$P(b_i \text{ wins}) = P\left(b_i < b_j, \forall j \neq i \wedge \frac{1}{1.05}b_i < b_k \forall k\right)$$
$$\Pi(b_i, c_i) = (b_i - c_i) \cdot P\left(b_i < b_j, \forall j \neq i \wedge b_i < 1.05 b_k, \forall k\right)$$

Probability all $n_S - 1, n_L$ costs are higher, as the bid function is monotonic

$$\Pi(b_i, c_i) = (b_i - c_i) \cdot \left(\left(1 - F_S(\beta_S^{-1}(b_i) \mid E)\right)^{n_S - 1} \left(1 - F_L(\beta_L^{-1}(\frac{1}{1.05}b_i) \mid E)\right)^{n_L}\right)$$

Similarly, for a large business i with bid $b_i = \beta_L(c_i)$, with small businesses j and other large busineses $k \neq i$:

$$P(b_i \text{ wins}) = P\left(b_i < \frac{1}{1.05} \cdot b_j, \forall j \wedge b_i < b_k, \forall k \neq i\right)$$
$$\Pi(\beta_L(c_i), c_i) = (b_i - c_i) \cdot P\left(b_i < \frac{1}{1.05} \cdot b_j, \forall j \wedge b_i < b_k, \forall k \neq i\right)$$
$$= (b_i - c_i) \cdot \left(\left(1 - F_S(\beta_S^{-1}(1.05 b_i) \mid E)\right)^{n_S} \left(1 - F_L(\beta_L^{-1}(b_i) \mid E)\right)^{n_L - 1}\right)$$

Next, we want to maximize this profit function with respect to the bid b_i, since we can manipulate the resulting first-order condition to obtain an equation for costs in terms of bids. Companies will choose their bid based on $\max_{b_i} \Pi(b_i, c_i)$. Note that for $y = m(x)$, we have

$$\frac{dm^{-1}(y)}{dy} = \frac{1}{m'(m^{-1}(y))}$$

For a small business, this maximization gives the following after applying the chain and product rule to the derivative:

$$\frac{\partial}{\partial b_i}(b_i - c_i) \cdot \left(\left(1 - F_S(\beta_S^{-1}(b_i) \mid E)\right)^{n_S - 1} \left(1 - F_L(\beta_L^{-1}(\frac{b_i}{1.05}) \mid E)\right)^{n_L}\right)$$

$$\implies (b_i - c_i) \cdot (\left(1 - F_L(\beta_L^{-1}(\frac{1}{1.05}b_i) \mid E)\right)^{n_L} \cdot (n_S - 1)(1 - F_S(\beta_S^{-1}(b_i)))^{n_S - 2} \cdot \frac{-f_S(\beta_S^{-1}(b_i))}{\beta_S'(\beta_S^{-1}(b_i))}$$
$$+ \left(1 - F_S(\beta_S^{-1}(b_i) \mid E)\right)^{n_S - 1} \cdot n_L(1 - F_L(\beta_L^{-1}(\frac{1}{1.05}b_i) \mid E))^{n_L - 1}) \cdot \text{ (cont. next line)}$$
$$(-f_L(\beta_L^{-1}(\frac{1}{1.05}b_i) \mid E)) \cdot \frac{1}{1.05\beta_L'(\beta_L^{-1}(\frac{b_i}{1.05}))})$$
$$+ \left(1 - F_S(\beta_S^{-1}(b_i) \mid E)\right)^{n_S - 1} \left(1 - F_L(\beta_L^{-1}(\frac{1}{1.05}b_i) \mid E)\right)^{n_L}) = 0$$

We can then move the last term over and divide through by it, giving

$$-1 = (b_i - c_i)((n_S - 1)(1 - F_S(\beta_S^{-1}(b_i)))^{-1}(-f_S(\beta_S^{-1}(b_i))) \cdot \frac{1}{\beta_S'(\beta_S^{-1}(b_i))} +$$
$$+ n_L(1 - F_L(\beta_L^{-1}(\frac{b_i}{1.05})))^{-1}(-f_L(\beta_L^{-1}(\frac{1}{1.05}b_i) \mid E)) \cdot \frac{1}{1.05\beta_L'(\beta_L^{-1}(\frac{b_i}{1.05}))})$$

Taking away a -1 on each side and substituting $\beta_S^{-1}(b_i)$ with c_i gives the differential equation

$$1 = (b_i - c_i) \cdot \left(\frac{(n_S - 1) \cdot f_S(c_i \mid E)}{1 - F_S(c_i \mid E) \cdot \beta_S'(c_i)} + \frac{(n_L) \cdot f_L(\frac{c_i}{1.05} \mid E)}{1 - F_L(\frac{c_i}{1.05} \mid E) \cdot \beta_L'(c_i)}\right)$$

Following the same process for a large business, the first order condition is equivalent to:

$$1 = (b_i - c_i) \cdot \left(\frac{(n_S) \cdot f_S(1.05c_i \mid E)}{1 - F_S(1.05c_i \mid E) \cdot \beta_S'(c_i)} + \frac{(n_L - 1) \cdot f_L(c_i \mid E)}{1 - F_L(c_i \mid E) \cdot \beta_L'(c_i)} \right)$$

Together we have a system of equations satisfying $\beta_S(\bar{c}) = 1.05\beta_L(\bar{c}) = \bar{c}$ and $\beta_S(\underline{c}) = 1.05\beta_L(\underline{c})$. From this system we will be able to obtain the costs and $F_{L \text{ or } S}(\cdot \mid E)$.

4 Identification and Empirical Estimation Strategy

4.1 Cost Distribution

While we have all the submitted bids, we do not have the original costs businesses estimated for themselves. However, assuming all bidders used an optimal strategy, we can recover those costs.

Recall the FOCs and differential equations:

$$\max_{b_i} \Pi(b_i, c_i), \qquad \max_{b_i} \Pi(b_i, c_i)$$

$$1 = (b_i - c_i) \cdot \left(\frac{(n_S - 1) \cdot f_S(c_i \mid E)}{1 - F_S(c_i \mid E) \cdot \beta_S'(c_i)} + \frac{(n_L) \cdot f_L(1.05c_i \mid E)}{1 - F_L(1.05c_i \mid E) \cdot \beta_L'(c_i)} \right)$$

$$1 = (b_i - c_i) \cdot \left(\frac{(n_S) \cdot f_S(\frac{1}{1.05}c_i \mid E)}{1 - F_S(\frac{c_i}{1.05} \mid E) \cdot \beta_S'(c_i)} + \frac{(n_L - 1) \cdot f_L(c_i \mid E)}{1 - F_L(c_i \mid E) \cdot \beta_L'(c_i)} \right)$$

These have the boundary conditions $\beta_S(\bar{c}) = 1.05\beta_L(\bar{c}) = 1.05\bar{c}$ and $\beta_S(\underline{c}) = 1.05\beta_L(\underline{c})$. Just as f is the pdf of costs, denote the pdf of bids by g.

Now recall that $G(b; E, n_S, n_L$ is the CDF of bids (with parameters: engineer's estimate, number of small businesses, and number of large businesses) and $F(c; E)$ is the CDF of costs, with costs $c_i = \beta^{-1}(b_i)$ for the appropriate β_S, β_L. Note that the bids depend on number of bidders, but not the costs; both are dependent on estimates, so everything we show below is conditional on E. Furthermore, bidding functions are strictly monotonic and therefore one-to-one (map one cost uniquely to one bid). Given this, a lower bid implies a firm also has a lower cost. We show that for one bidder's bid B:

$$\begin{aligned} 1 - G_S(b_i \mid n_S, n_L, E) &= P(\beta_S^{-1}(B) \geq \beta_S^{-1}(b) \mid E) \text{ because bids are monotonic} \\ &= P(C \geq B_S^{-1}(b) \mid E) = 1 - F_S(\beta_S^{-1}(b); E) \\ G_S(b_i \mid n_S, n_L) &= F_S(\beta_S^{-1}(b_i); E) \end{aligned}$$

and

$$\begin{aligned} 1 - G_L(b_i \mid n_S, n_L, E) &= P(\beta_L^{-1}(B) \geq \beta_L^{-1}(b) \mid E) \text{ because bids are monotonic} \\ &= P(C \geq B_L^{-1}(b)) = 1 - F_L(\beta_L^{-1}(b); E) \\ G_L(b_i \mid n_S, n_L) &= F_L(\beta_L^{-1}(b_i); E). \end{aligned}$$

Furthermore, to get g_S and g_L we must take the derivatives of G_S and G_L: $\frac{dx}{dy} = \frac{1}{m'(x)} \implies \frac{dm^{-1}}{dy} = \frac{1}{m'(m^{-1}(y))}$:

$$\frac{d}{db_i} G_L(b_i \mid n_S, n_L, E) = \frac{d}{db_i} F_L(\beta_L^{-1}(b_i); E)$$
$$g_L(b_i \mid n_S, n_L, E) = f_L(\beta_L^{-1}(\beta_L(c_i)))\beta_L'(\beta_L^{-1}(b_i))$$

Similarly,

$$F_S'(\beta_S^{-1}(b); E) = f_S(\beta_S^{-1}(b)) \cdot \frac{1}{\beta_S'(\beta_S^{-1}(b))}$$

4.1.1 Recovering Bidder Costs

Plugging this in to the partial differential equation which gives the optimal bidding function for small businesses, we obtain:

$$1 = (b_i - c_i) \cdot \left(\frac{(n_S - 1) \cdot g_S(\beta_S(c_i \mid E) \mid n_S, n_L) \cdot \beta_S'(\beta_1^{-1}(b)}{1 - G_S(b_i \mid n_1, n_2) \cdot \beta_S'(c_i)} + \frac{(n_L) \cdot g_L(1.05 b_i \mid n_L, n_S) \cdot \beta_L'(c_i)}{1 - G_L(1.05 b_i \mid n_L, n_S) \cdot \beta_L'(c_i)} \right)$$
$$= (b_i - c_i) \cdot \left(\frac{(n_S - 1) \cdot g_S(\beta_S(c_i \mid E) \mid n_S, n_L)}{1 - G_S(b_i \mid n_1, n_2) \cdot} + \frac{(n_L) \cdot g_L(1.05 b_i \mid n_L, n_S)}{1 - G_L(1.05 b_i \mid n_L, n_S)} \right)$$
$$c_i(b_i; E, n_S, n_L) = b_i - \left(\frac{(n_S - 1) \cdot g_S(\beta_S(c_i \mid E) \mid n_S, n_L)}{1 - G_S(b_i \mid n_1, n_2) \cdot} + \frac{(n_L) \cdot g_L(1.05 b_i \mid n_L, n_S)}{1 - G_L(1.05 b_i \mid n_L, n_S)} \right)^{-1}$$

Similarly, we can put the costs in terms of bids and g_L for large businesses:

$$1 = (b_i - c_i) \cdot \left(\frac{(n_S) \cdot g_S(\frac{1}{1.05} \beta_S(c_i \mid E) \mid n_S, n_L) \cdot \beta_S'(\beta_1^{-1}(b)}{1 - G_S(\frac{1}{1.05} b_i \mid n_1, n_2) \cdot \beta_S'(c_i)} + \frac{(n_L - 1) \cdot g_L(b_i \mid n_L, n_S) \cdot \beta_L'(c_i)}{1 - G_L(b_i \mid n_L, n_S) \cdot \beta_L'(c_i)} \right)$$
$$= (b_i - c_i) \cdot \left(\frac{(n_S) \cdot g_S(\frac{1}{1.05} \beta_S(c_i \mid E) \mid n_S, n_L)}{1 - G_S(\frac{1}{1.05} b_i \mid n_1, n_2) \cdot} + \frac{(n_L - 1) \cdot g_L(b_i \mid n_L, n_S)}{1 - G_L(b_i \mid n_L, n_S)} \right)$$
$$c_i(b_i; E, n_S, n_L) = b_i - \left(\frac{(n_S) \cdot g_S(\frac{1}{1.05} b_i \mid n_S, n_L)}{1 - G_S(\frac{1}{1.05} b_i \mid n_1, n_2) \cdot} + \frac{(n_L - 1) \cdot g_L(b_i \mid n_L, n_S)}{1 - G_L(b_i \mid n_L, n_S)} \right)^{-1}$$

In the next section, we can then use the bid data and kernel density estimates of G_S, G_L to recover estimates of the cost.

5 Estimation Results

The goal of this section was to recover estimated costs for each bidder.

To simplify the conditional statements that we will have to make, we will only focus on the most common auctions that have a particular pair of number of small and large businesses. The largest sample in the Cal Trans data set is one small business and three large businesses. Focusing on only one pair of n_S and n_L simplifies the calculation of density functions.

As we found in the previous section, for small businesses:

$$c_i(b_i; E, n_S, n_L) = b_i - \left(\frac{(n_S - 1) \cdot g_S(\beta_S(c_i \mid E) \mid n_S, n_L)}{1 - G_S(b_i \mid n_1, n_2) \cdot} + \frac{(n_L) \cdot g_L(1.05 b_i \mid n_L, n_S)}{1 - G_L(1.05 b_i \mid n_L, n_S)} \right)^{-1}$$

and for large businesses:

$$c_i(b_i; E, n_S, n_L) = b_i - \left(\frac{(n_S) \cdot g_S(\frac{1}{1.05} b_i \mid n_S, n_L)}{1 - G_S(\frac{1}{1.05} b_i \mid n_1, n_2) \cdot} + \frac{(n_L - 1) \cdot g_L(b_i \mid n_L, n_S)}{1 - G_L(b_i \mid n_L, n_S)} \right)^{-1}$$

In these equations, g_s, G_s, etc., represent the pdf and cdf of the bids. To find them, we use the definition of conditional probability:

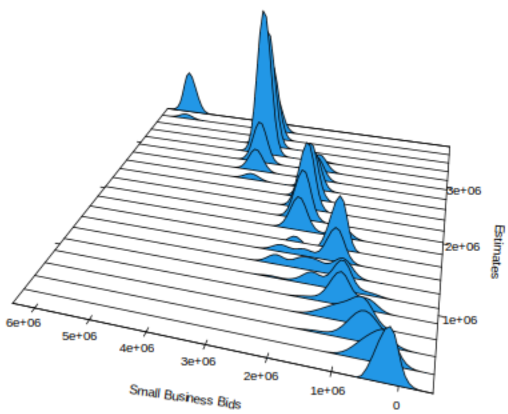
$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)}$$

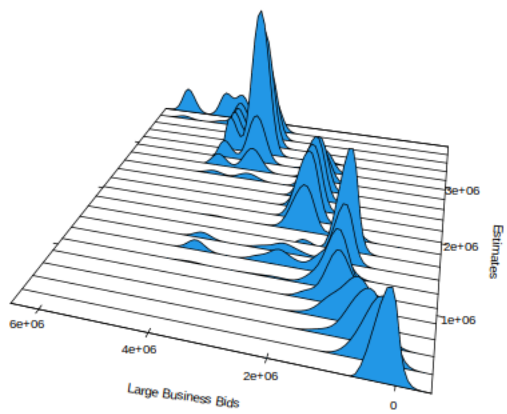
Therefore, we need to estimate densities for the joint probability of bids and estimates and the marginal probability of estimates.

We will first estimate functions for the bids (i.e., g_s and g_l in the above equation). They will be conditional on the engineers estimate. To simplify this, we will condition on the median of the engineers estimate. For our subset of the auctions with the most common combination of large/small bidders, this was \$526000.

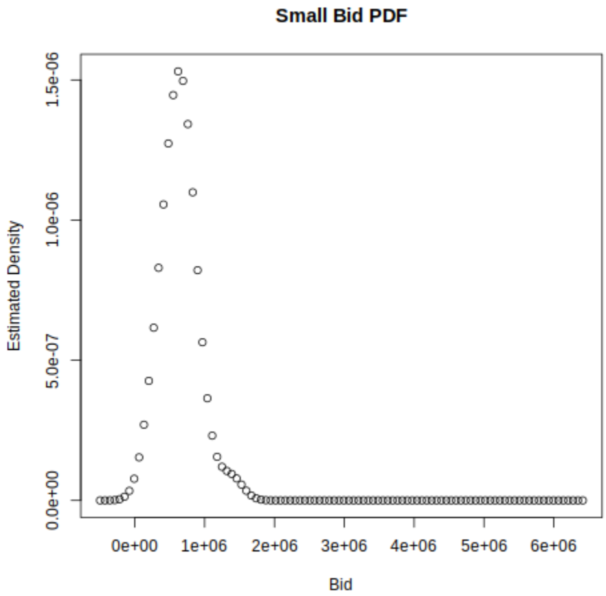
We will now estimate the PDF of small business bids which is conditional on the median of the engineers estimate. It will not be conditional on the number of small and large businesses because we are fixing it to our chosen subset of auctions.

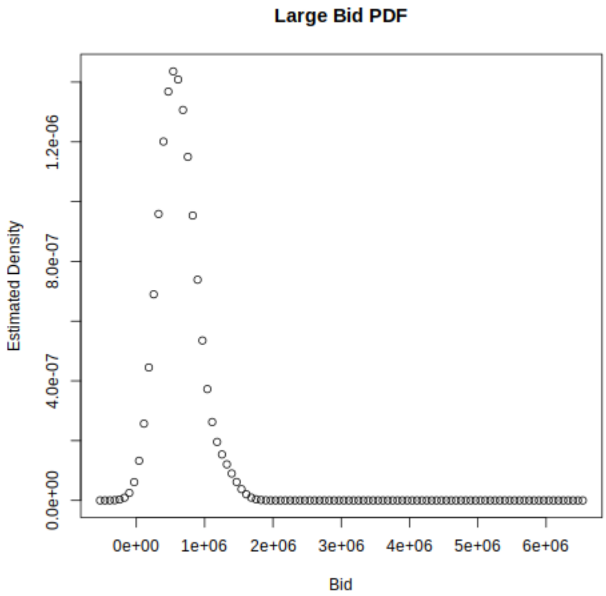
First, here are 3D plots of the conditional PDFs of bids on estimates, for small and large businesses.





Then, here are the graphs of the PDF for small/large businesses evaluated on their bids, conditional on the median estimate from our subset (which is the same for both large and small businesses, given they compete in the same auctions).

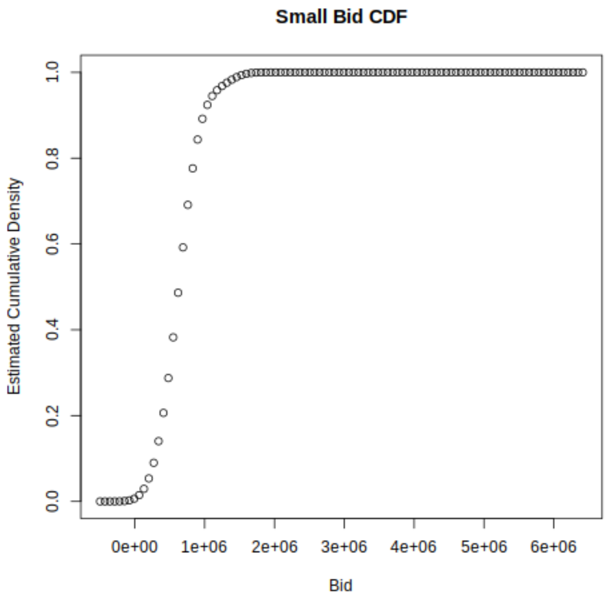


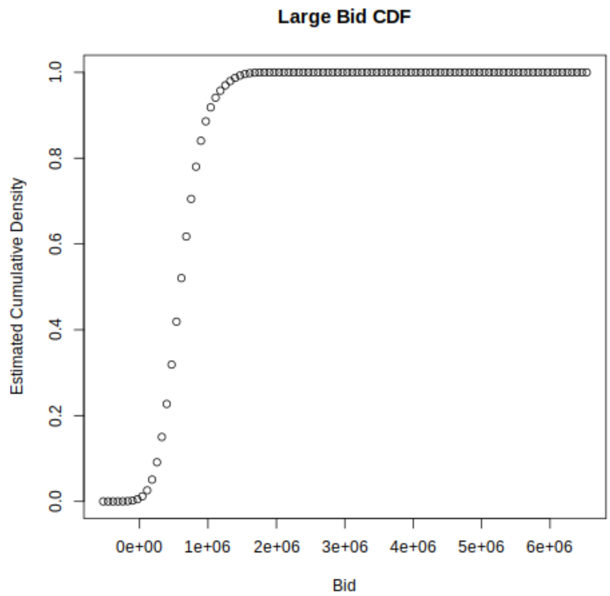


While it is counterintuitive that the small businesses actually have a mode slightly to the right of the large businesses, in our subset of the data the median small business bid was \$615605 and the median large business bid was \$585513. The large businesses do have a handful of bids on much larger projects (up to 50 million dollars in the full data set), but they are not present in our subset of the data.

Even within our subset, the number of large business bids on larger projects is small enough to not be very visible on the density plots. However, once we recover the costs, it becomes more noticeable.

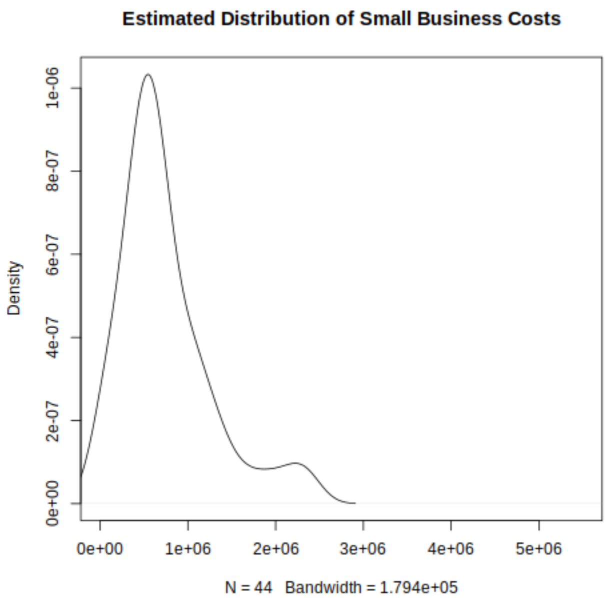
Similarly, below are the two CDFs of bids ($\boldsymbol{G_s}$ and $\boldsymbol{G_l}$ in the above equation).

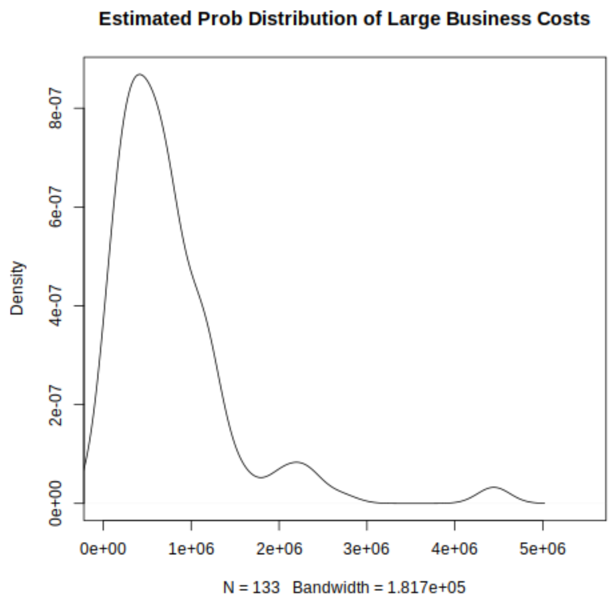


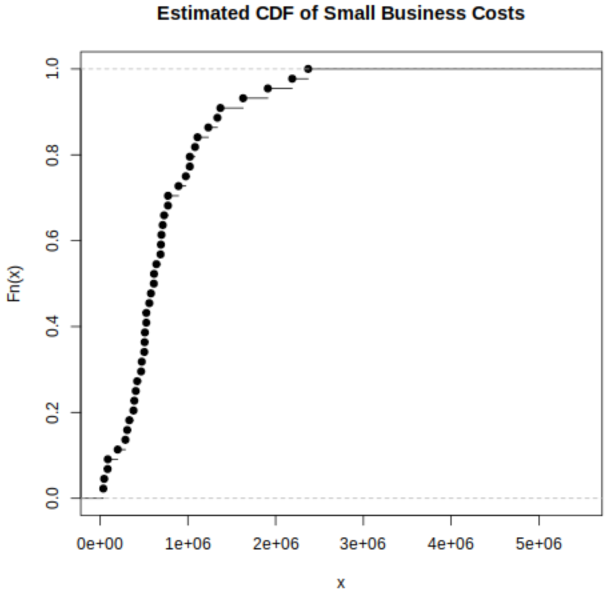


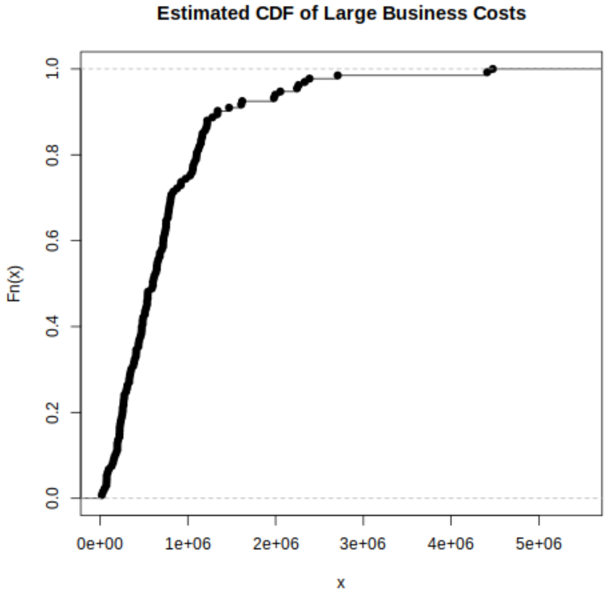
Finally, we calculate the estimated cost distribution by recovering the costs from the bid distribution, using the equation above (calculated in the Identification section).

The scale of these plots was set to \$6 million for both small and large businesses: the reason the small business plots stop much earlier is because they generally did not participate in auctions for the very large projects, but rather in auctions where their costs would be smaller (typically under a million dollars).









6 Counterfactual

To favor small businesses, instead of using the 5% rule which exhibits a preference towards small business auction participants, Caltrans could implement reserve prices in their auctions. A reserve price is a price set by the auctioneer such that if all bidders bid above the reserve price, no one will win (for low-bid auction). For Caltrans, if the reserve price is not met, then the project would cost the engineer's estimate and Caltrans would carry out the project instead of outsourcing it. We will determine two reserve prices for Caltrans auctions: one for small businesses and one for large businesses. These auctions with reserve prices will be a second price auction, such that the project will cost the second lowest bid. We will use this to study a theoretical alternative where CalTrans had performed second price auction with reserve prices, using costs because it is optimal to bid one's cost in SPA.

The assignment of the auction winner and cost follows the following rules. If a small business has the lowest bid, but they do not meet the reserve price, then the project will cost the engineers estimate and nobody wins the auction. Next, if a small business wins the auction and bids below or at the reserve price, then if the second lowest bid is below the reserve price, the winner will get paid the second lowest bid. However, if the second lowest bid is above the reserve price, the winner will get paid the reserve price. Another scenario is that a large business wins the auction, but does not bid below the reserve price. If the small business is the second lowest bid and is below their reserve price, then they will win the auction and pay the reserve price for small businesses. Finally, if a large business wins the auction and bids below their reserve price, then they will carry out the project and get paid the second lowest bid if it is below their reserve price, or the reserve price for large businesses if the second lowest bid is above that reserve price.

To calculate the reserve prices for small and large businesses we used the following equation:

$$r* = c_0 - \frac{F(r*)}{f(r*)}$$

Note that as long as small businesses tend to have higher costs, a 'tiered' reserve price system set this way will favor them.

The optimal reserve price is r*. We used F_s and f_s for small businesses and F_l and f_l for large businesses, while using the median engineer's estimate (to make small and large businesses comparable). We will also denote the small business reserve prices by r_s and r_l for the large business reserve prices. We want to determine how this new rule affects costs for Caltrans. To determine this we randomly draw 1 cost from F_s and 3 costs from F_l (since as before, we are focusing on the largest subset of auctions with those n_L and n_S for simplicity). This is repeated 1,000 times to find an Monte Carlo average. We found the average cost for SPA with reserve price to be around \$390,000, but it is important to note that the average cost is different each time. Additionally, R was intermittently unable to find costs from our estimated CDF via inverse transform sampling at the tail end due to floating point errors in the uniroot function. For this reason, we are only sampling from the 0.01 to 0.90th percentiles of the distribution. It is thus likely that the average cost with reserve price is slightly underestimated.

To find the average cost of the auction format with the 5% rule, we determined the winning bids of all the auctions from our subset and then took the average. This average was \$726,000, and is a known empirical result from the historical Caltrans data.

The difference between the average cost from SPA with reserve prices and the average cost from FPA with bidder preference aligns with what auction theory tells us. Without reserve prices the revenue from SPA and FPA should be equal due to the Revenue Equivalence Theorem as long as the assignment rules are exactly the same. However, our optimal reserve prices will not necessarily always lead to the same results as the 5% rule, since they do not follow that standard but instead minimize costs for Caltrans.

When revenue equivalence holds and once a reserve price is added, expected revenue should increase. Despite the strongest form of the revenue equivalence theorem not holding here, revenue should still be similar between first and second price auction. The marked difference in expected costs that we found in our simulations, where the average cost with a reserve price is considerably lower compared to the average cost with only bidder preference, reflects this result. A lower cost is better for Caltrans because they do not have to spend as much on their highway projects. Further research on the effect of reserve prices on public auctions should be undertaken to confirm this.

7 Conclusion

We were able to obtain estimates of the distribution of costs (analagous to private values) for bidders on CalTrans procurement projects using data on their bids and estimates of the true cost. This is because there is an optimal bidding function which we assume the bidders followed. The work was complicated by the existence of small- and large-business type bidders.

The first step of this analysis was to explore bidding behavior by modeling the bidding functions for small business bidders and large business bidders. These models demonstrated the probability of winning a particular auction, since in first price procurement auction, the bidder with the lowest cost or value will win the auction, assuming all participants are rational. We then maximized the expected profits using optimization to recover the costs, as the Caltrans data only reports submitted bids. Finally, using conditional kernel density estimation, we found estimated cost distributions for small business bidders and large business bidders.

Using a subset of the bidding data, as well as the data on engineers' estimates, we plotted the PDFs and CDFs of the bidding distributions and cost distributions for both small business bidders and large business bidders. To simplify these calculations we only focused on auctions with the most prominent combination of small business and large business bidders. This ends up being auctions with one small business and three large businesses. These plots were our main results and demonstrated the respective behaviors of such bidders.

One addendum that we explored with this data was a counterfactual scenario in which second price auction with reserve prices was used. This explored what costs may have been for Caltrans if there was no bidder preference or '5% rule.' Instead, if Caltrans had set specific reserve prices for small businesses and large businesses, we found that the average cost of Caltrans projects would have been lower. Thus, reserve prices may be beneficial for Caltrans and other public utilities because they decrease their costs.

This paper exhibits three main weaknesses. The first was the simplification to only look at auctions with 1 small and 3 large business bidders, which notably reduced the sample size from 705 to 59; and the use of the median engineer's estimate throughout. The second was the reliance on kernel density estimation to model bidding behavior, which is only asymptotically unbiased and presented an issue as those density estimates were plugged in to recover cost estimates. The last was issues with floating point error when dealing with the inverse of very large bids, which we could see with inverse transform sampling in the counterfactual section and with a handful of 'infinite' estimates when recovering costs, which had to be included.

Overall, the findings of this report demonstrated how firms bid in these Caltrans low-bid procurement auctions. Using auction theory principles, we are able to evaluate and explore bidding behavior in real auctions. These findings allow us to predict in future auctions how opposing bidders will act, so that bidders can maximize their profits. Each case study gives us more information as to how bidders act in auctions.