On-line Appendix for "Procurement Contracting with Time Incentives: Theory and Evidence"

Gregory Lewis
Harvard University and NBER

Patrick Bajari
University of Minnesota and NBER

January 3, 2011

1 Introduction

In this supplementary appendix, we present additional evidence and robustness analysis on three main topics. First, we argue that our description of how Caltrans makes its decisions on A+B assignment and the usercosts associated with those contracts is consistent with the data. Second, we consider alternative treatment evaluation strategies based on nearest neighbor matching and regression discontinuity respectively, and show that the estimates we obtain are similar to those in the main text, albeit far less precisely estimated. Finally, we give a detailed description of the simulation procedure for the counterfactuals, and then test the sensitivity of our results by replicating our counterfactual analyses under a variety of alternative scenarios. This exercise leads us to conclude that the main qualitative themes of the paper are extremely robust: that the A+B policy as currently used has been successful, but that expanded use of time incentives would lead to significant welfare improvements, potentially at relatively low cost to Caltrans.

2 Caltrans Policy Rules

As we said in the main paper, Caltrans officials indicated two distinct policies regarding assignment and usercosts. District 4 assigned A+B status to long-duration contracts (those

estimated to last over 100 working days), while the remaining districts assigned them to big contracts (engineer's estimate over \$5M) that were projected to severely inconvenience commuters (daily welfare cost over \$5000). We call projects that are over 100 days in length or over \$5M in size, "eligible". Now, conditional on assigning a contract to be A+B, district 4 chooses the usercost based on the liquidated damage formula:

$$\text{Liquidated Damages} = \left(\text{LD\%} \times \frac{\text{Engineer's Estimate}}{\text{Engineer's Days Estimate}}\right) + \text{RE Office Expenses} \quad (1)$$

where the LD% depends on the type and size of contract, and ranges from 10% to 20%. The home office expenses are meant to capture the office expenses for the resident engineer while on the project (e.g. office rental). On the other hand, the remaining districts set the usercost equal to their estimate of the daily welfare cost.

In this section, we test these claims. First, we look at A+B assignment. Less than 1% of ineligible contracts are assigned to be A+B, whereas 41% of eligible contracts are assigned to be A+B. This is consistent with the rule indicated by Caltrans. To understand assignment within the pool of eligible contracts, we run a a probit for A+B assignment using a small set of RHS covariates, estimated only on the sample of size-eligible contracts. The results are shown in Table 1. Looking at the third column (for all contracts), we see that bigger contracts and those with more traffic are significantly more likely to be assigned A+B status. If we use a rule that predicts A+B assignment by assuming that ineligible contracts are never A+B (consistent with the policy) and eligible contracts are A+B only if their predicted probability from the probit is above 50%, we can correctly predict nearly 92% of the contract assignments. This is not bad: one benchmark is how many would be correctly predicted by assuming all contracts were standard, and then you'd only get 89%. This difference is amplified if one looks only at size-eligible contracts: 71% correctly predicted using the probit versus 59% if always predicted to be standard.

For the usercosts, the formula does a good job of predicting district 4 behavior. This is shown in Figure 1, which plots log usercost against log (engineer's estimate / engineer's days) for district 4 and the other districts (best fit lines are shown). Taking logs on both sides in (1), one would expect a roughly linear relationship in district 4 (ignoring the home office expenses), and this is basically what we see in the figure. This is supported by regression analysis. Table 2 shows the results of a regression of log usercost on covariates, where the log specification is chosen because of the ratio form of the liquidated damage formula. For

Table 1: Probability of A+B Assignment

	A+B Assignment		
	District 4	Other Districts	All
Log Engineer's Estimate	0.1688***	-0.0381	0.1250***
	(0.0357)	(0.0716)	(0.0340)
Log Engineer's Days	-0.1533	-0.0172	-0.1402**
	(0.0924)	(0.0859)	(0.0583)
Log Daily Traffic	0.0879**	0.0533	0.0806***
	(0.0360)	(0.0381)	(0.0272)
District 4			0.2779***
			(0.0593)
N	94	91	185
% Correctly Predicted (eligible only)	72.34	71.43	71.35
Benchmark Prediction % (eligible only)	53.19	71.43	58.92
% Correctly Predicted (all contracts)	85.50	94.69	91.95
Benchmark Prediction % (all contracts)	73.50	94.69	88.70

Coefficients are average marginal effects from a probit of AB assignment on the covariates using the sample of eligible contracts only. "Eligible" refers to contracts with engineer's days > 100 in District 4, and engineers estimate $>$ \$5M elsewhere. Reported standard errors are robust. Significance is denoted by asterisks (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). The % correctly predicted is the fraction of A+B assignment decisions predicted by a rule that only assigns AB status to eligible contracts with $\widehat{p} > 0.5$. Benchmark prediction percentages are from a rule that uniformly predicts A+B or standard assignment for all contracts, depending on which contract type is more prevalent in the relevant sample.

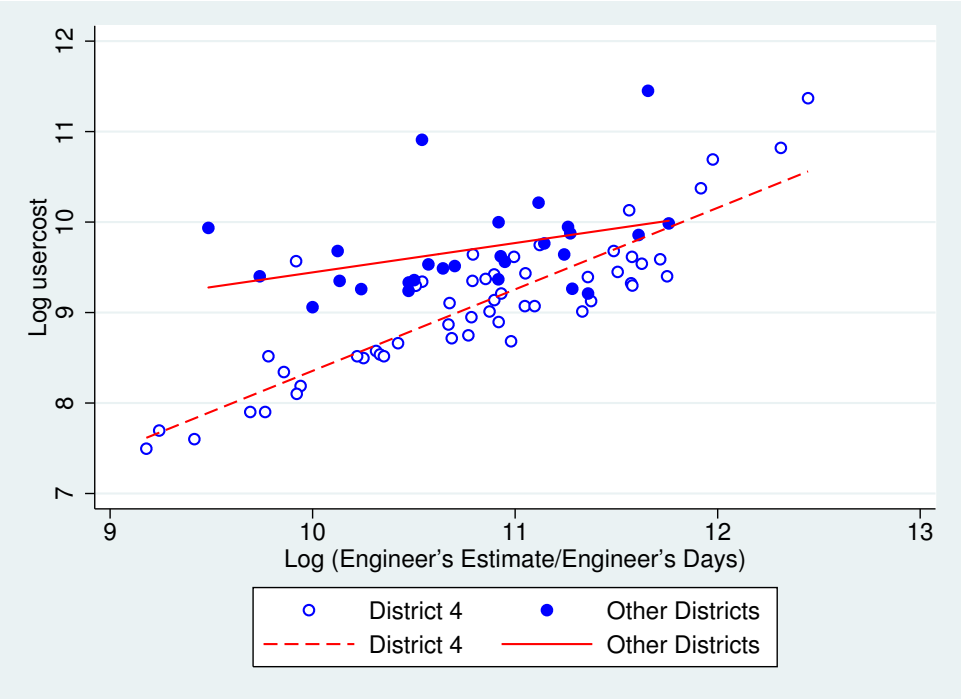

Figure 1: **Assignment of Usercosts.** Shown is a scatterplot of the log usercost against log (engineer's estimate / engineer's days) for district 4 and the other districts, with best fit lines superimposed over it. For district 4, the relationship is close to linear, consistent with the formula district 4 say they use.

district 4, we can explain 80% of the variation, again suggesting that their description is accurate. We are less successful for the other districts (R^2 of 0.42), but this is probably not surprising given the limited amount of data and the variation in methodology for calculating social costs across districts. All in all, the data is consistent with the policy descriptions given to us by Caltrans, although there is certainly noise in both the assignment and the usercost choices.

3 Alternate Policy Evaluation Strategies

In the main text, we estimated the average treatment effect on the treated (ATT) of A+B assignment, for a variety of dependent variables. For example, we estimated the ATT on the winning bid. To do this, we looked at the difference between log winning bids between the treatment group (A+B contracts) and a control group (standard contracts that were A+B eligible), after showing that the treatment and control group were balanced on most

Table 2: Predicting Usercosts in A+B Contracts

	Log Usercost		
	District 4	Other Districts	All
Log Engineer's Estimate	0.952***	0.374**	0.790***
	(0.063)	(0.166)	(0.081)
Log Engineer's Days	-0.928***	-0.052	-0.674***
	(0.091)	(0.239)	(0.121)
Log Daily Traffic	0.027	0.014	0.070
	(0.041)	(0.082)	(0.052)
Plant Establishment Work	0.348**	0.229	0.244**
	(0.169)	(0.163)	(0.113)
Log Liquidated Damage %	0.704***	-0.096	0.462*
	(0.250)	(0.668)	(0.275)
District 4			-0.644***
			(0.110)
N	53	27	80
R^2	0.83	0.23	0.71

Robust standard errors reported. Significance is denoted by asterisks (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

other observables. To implement this, we used an OLS regression with many covariates and a dummy for A+B assignment.

Although simple, this approach faces two main critiques. The first is that the treatment effect may be non-linear, and so imposing a linear or log-linear functional form as in the OLS regressions may introduce bias. The second is that it assumes no selection on unobservables. In this section we try two alternative approaches designed to correct for these two problems in turn: nearest neighbor matching and regression discontinuity design.

3.1 Nearest Neighbor Matching

The idea of nearest neighbor matching is to match each treated observation with a fixed number of control observations that are most similar on the observables, and measure the (average) difference in the dependent variable. Averaging these differences gives an estimate of the ATT. We implement this by matching on log engineer's estimate, log engineer's days, log traffic, Lane Closure Fraction, a dummy for re-opening penalty, year, type of work and district fixed effects. Each treatment observation is matched with 3 control observations. We

use the linear bias-correction of Abadie, Drukker, Herr and Imbens (2004), since even small differences in average contract size across the treatment and control might otherwise bias the results. We force "exact" matching on a dummy for district 4, implying that every treatment observation in district 4 is matched with control observations in district 4. Exact matches are achieved in 100% of cases. The estimates are sensitive to these choices (how many matches, or the exact matching variables), and so the estimates shown below are somewhat artificially precise, particularly in the case of the winning bid (the ATT's on completion time are quite robust).

Table 3 shows the estimated results. They are remarkably similar to those in the main text: when estimated off the full sample A+B contracts cost around 6.5-7.5% more, and between 5.7% (OLS) and 9.5% (matching) off the subset of completed contracts. Contracts are estimated in both cases to take about 32% less working days when they are A+B, or 35% (OLS) to 38% (matching) fewer total days (including weather and other days). The ATT for quality deductions is statistically and economically insignificant in both cases (the interpretation of the coefficient shown below is that A+B assignment raises quality deductions by 0.0036% of the engineer's estimate). Overall, the results seem similar to those we obtained by OLS, limiting the concern of bias introduced by functional form assumptions.

Table 3: Alternative Policy Evaluation by Nearest Neighbor Matching

	Eligible contracts		Eligible & Completed		
	(1)	(2)	(3)	(4)	(5)
ATT	0.067**	0.095*	-0.317***	-0.373***	0.036
	(0.034)	(0.049)	(0.041)	(0.136)	(0.177)

In all columns, the average treatment on the treated (ATT) is estimated by nearest neighbor matching. Column (1) is estimated off all A+B eligible contracts, and the dependent variable is log winning bid. In columns (2)-(5), the estimation sample is all completed and A+B eligible contracts, and the dependent variables are log winning bid, working days / engineer's days, total days / engineer's days, and quality deductions $\times$ 1000 / engineer's estimate respectively. Matching is on the basis of log engineer's estimate, log engineer's days, log traffic, Lane Closure Fraction, a dummy for re-opening penalty, year, type of work and district fixed effects. Exact matching on a dummy for district 4 is employed, and each treatment observation is matched with 3 control observations. Exact matches are achieved in 100% of cases. Significance levels are denoted by asterisks (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

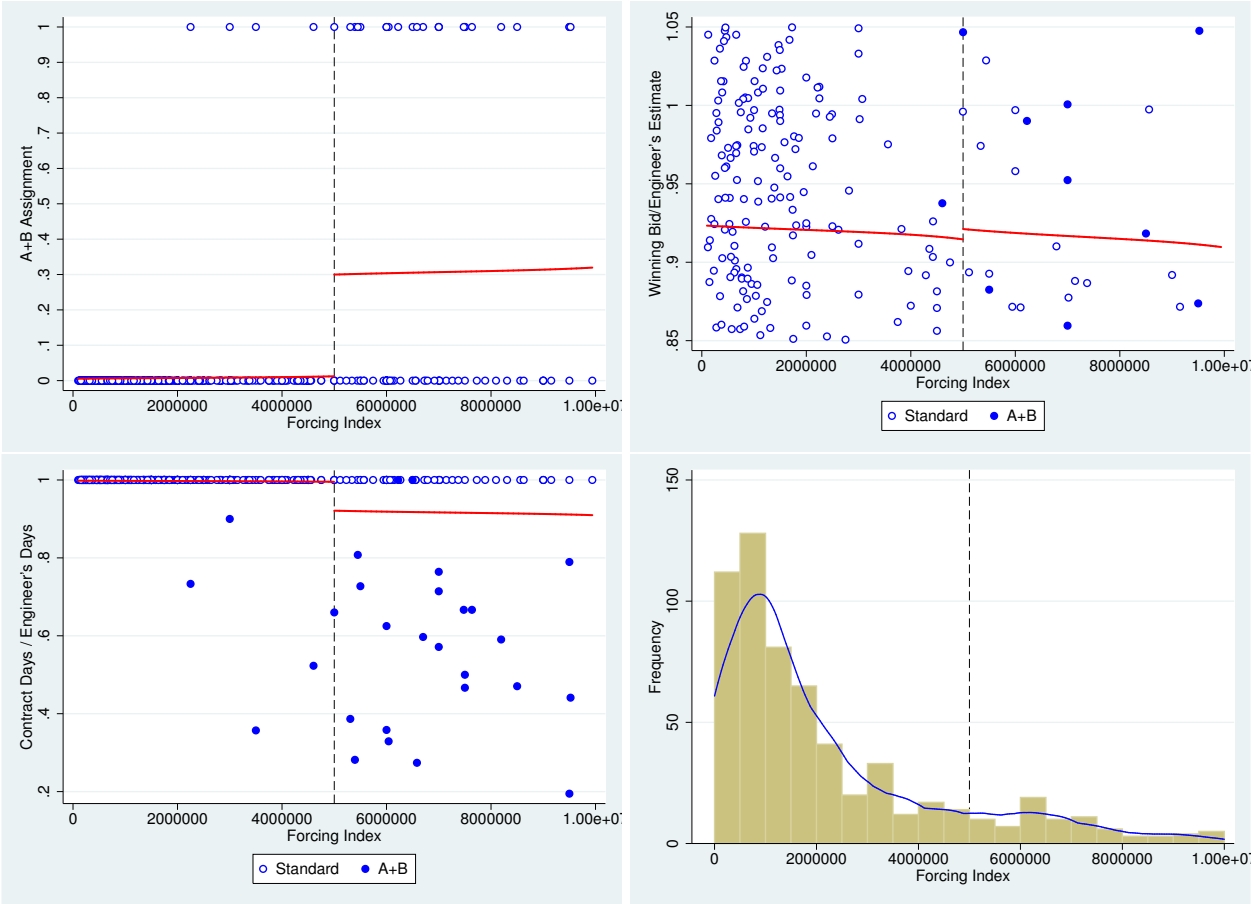

Figure 2: **RDD Approach.** The four graphs illustrate the RDD approach. The forcing index is defined as the engineer's estimate outside of district 4, and as 5000 × the engineer's days estimate in district 4. According to Caltrans policy, A+B eligibility occurs when the forcing index exceeds 5 million. The top left panel is the change in A+B assignment around this threshold; the top right panel shows the change in winning bid / engineer's estimate around that point; the bottom left panel shows the change in the ratio of contract days to engineer's days there; and the bottom right panel is the density of the forcing index around the threshold. All lines are based on local polynomial regressions with bandwidth \$2.5M (except for the density, bandwidth \$0.5M).

3.2 Regression Discontinuity Design

The strategy here is to make use of the policy rule that we examined earlier, and to compare contracts that are just below the cutoff for size eligibility with those just above. Assuming that both the forcing variable and the unobservables have continuous density across the cutoff, this approach has the advantage of being robust to selection on unobservables (although it gives us a local treatment on the treated (LATT) rather than the average treatment on the treated (ATT) from earlier). The downside is that we simply don't have enough observations

very close to the cutoff to implement this convincingly, and so "local" here will be contracts of between \$2.5M and \$7.5M outside of district 4 (\$2.5M either side of the \$5M threshold), and between 50 and 150 days in district 4 (50 days either side of the 100 day threshold). We combine the analysis for both groups of contracts by defining a composite forcing index equal to the engineer's estimate outside of district 4, and as 5000 $\times$ the engineer's days estimate in district 4. We use the full sample of contracts throughout (in principle, we could do this only for completed contracts, but data limitations make this impractical).

Figure 2 shows all the relevant data.[1] The top left panel shows the change the probability of A+B assignment around the relevant threshold, as estimated by the gap in the local polynomial estimate of assignment probability to the left and to the right of the threshold. It is pretty clear that there is a discrete jump in this probability, which makes a fuzzy RDD design possible. The top right panel shows that there is also a small discrete jump in the ratio of winning bid to engineer's estimate at that point, around 1%. The bottom left panel shows that there is a substantive drop in the ratio of contract days to engineer's days. To the left of the threshold it is very close to 1, reflecting the fact that most contracts are standard and standard contracts have contract days equal to the engineer's days. To the right of the threshold is is closer to 0.9, since there are far more A+B contracts on that side. In all of these first three panels, the local polynomial estimate is computed using a Nadaraya-Watson kernel estimator with bandwidth 2.5 million and Epanechnikov kernel. The last panel, on the bottom right, shows that the density of the forcing variable appears to be relatively smooth across the threshold, which limits concerns that there may have been manipulation of the forcing variable.

We run regressions that replicate the graphical analysis, regressing some dependent variable on a dummy for the threshold, and adding district, year and type of work fixed effects. Observations are weighed according to their distance from the threshold using an Epanechnikov kernel with bandwidth 2.5 million. The results are shown in table 4. Crossing the threshold is estimated to increases the probability of A+B assignment by 24.1%, and to decrease the contract to engineer's day ratio by 10%. Both of these are significant. By rescaling, this implies a LATT on the day ratio of -41.6%, which is reassuringly close to our ATT estimate of -39.5% in the main paper. On the other hand we can't really pin down a cost increase by this method. Our point estimate on the increase in winning bid ratio across the threshold is 3%, implying a LATT of 12.6% with a standard error of around the same magnitude (versus

[1] In deciding which graphs were useful, we were guided by Imbens and Lemieux (2008).

Table 4: Alternative Policy Evaluation by Regression Discontinuity

	Coefficient	Standard Error
ITT A+B Assignment	0.241***	0.050
ITT Winning Bid / Engineer's Estimate	0.030	0.032
ITT Contract Days / Engineer's Days	-0.100***	0.024
LATT Winning Bid / Engineer's Estimate	0.126	0.136
LATT Contract Days / Engineer's Days	-0.416***	0.055

Results from a fuzzy RDD estimate of the effects of A+B assignment on winning bid / engineer's estimate and contract days / engineer's days. The forcing variable is defined as the engineer's estimate outside of district 4, and as 5000 $\times$ the engineer's days estimate in district 4. A+B eligibility occurs when the forcing variable exceeds 5 million, indicated by a dummy variable. The first row gives the ITT on A+B assignment, the second row gives the ITT on contract days/engineer's days, and the third row gives the ITT on winning bid/engineer's estimate. These are estimated by regressing the relevant dependent variables on the threshold dummy plus year, district and type of work fixed effects. Observations are weighted according to their distance from the threshold using an Epanechnikov kernel with bandwidth 2.5 million. The first and second ITT's are rescaled by the third to give a LATT for bids and days in rows 4 and 5 respectively. Standard errors are calculated via the delta method. Significance levels are denoted by asterisks (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

a 7.5% ATT in the main text). Moreover this estimate is sensitive to the choice of bandwidth and kernel, suggesting that the standard error in some ways understates our uncertainty. In most alternative specifications we have tried the LATT is positive, as one would expect, but we can't say too much more.

4 Counterfactuals

4.1 Simulation Procedure

The simulation procedure we use for the counterfactuals is as follows. Fix a counterfactual policy (i.e. a list of A+B contracts and associated usercosts). Then for each contract, randomly draw a set of participants using the estimated logit model, and independently draw contract shocks ξ_j and types θ_{ij}^A for each participant from the estimated normal distributions.[2] Then using the bidding model, we determine the level of acceleration that equates marginal

[2]Whenever the entry model randomly draws zero entrants, we force the draw of a single bidder, where each firm has chance proportional to their probability of participation.

benefits and costs for each participant, and work out the associated total acceleration costs.[3] Finally, we simulate bids (up to a contract fixed effect) for each contractor using the bidding model, and then deduce the winner of the auction according to the scoring rule.

For a given contract, we store five main variables:

- $\widetilde{d}_{Wj}$, the acceleration offered by the winning contractor (hence the index W)
- $c_A(\widetilde{d}_{Wj}, \theta^A_{Wj})$, the acceleration cost of the winning contractor;
- b_{Wj}, the bid of the winning contractor
- $\min_{i \in \mathcal{I}} b_{ij}$, the lowest dollar bid
- $\max_{i \in \mathcal{I}} c_A(\widetilde{d}_{ij}, \theta^A_{ij})$, the maximum acceleration cost across all bidders

Together these allow us to produce all the counterfactual welfare estimates (see below for the argument as to why this allows us to bound base costs). We repeat this simulation process 1000 times, and report the average outcomes across all simulations.[4] To get confidence intervals for the outcomes, we estimate the acceleration and bidding regressions, and recover type distributions exactly as before using a sequence of 100 bootstrap samples and then run the simulation using the bootstrapped coefficients.[5] We choose not to bootstrap the participation model since it is quite precisely estimated off a very large sample and estimation is computationally demanding. The bootstrap samples are clustered by contract and stratified by district, so that the fixed effects are estimable in all samples. We report 95% confidence intervals from the 2.5% and 97.5% quantiles of the bootstrap simulations.

4.2 Bounds on Changes in Base Costs

We would like to know something about the difference between the base cost of the winning contractor $c_B(\mathbf{x}_{ij}, \theta^B_{Wj})$ and the bidder with the lowest base cost $\min_{i \in \mathcal{I}} c_B(\mathbf{x}_{ij}, \theta^B_{ij})$. This tells us how much more expensive the winning contractor is, in terms of base costs, than the contractor who would have won a standard auction (assuming an efficient allocation in the

[3] We constrain the days bid to be no less than 20% of the engineer's days — essentially assuming infinite marginal costs at that point — since we never observe bids below this threshold in the data.

[4] We have experimented with more or less repetitions, but there is little simulation error at this size.

[5] We throw out bootstrap samples where the coefficient on log usercost is negative, as the simulation results become meaningless. This occurs in around 9% of cases.

standard auction; in this sense, it's an upper bound). Under a precise assumption on the markups, we can bound this difference.

To keep things simple, we drop most of the notation. For any bidder i, let $(b_i, \widetilde{d_i})$ be their bid, let c_i^T be the associated total costs for them of fulfilling that contract, broken down into base costs c_i^B and acceleration costs c_i^A. Also define their markup $m_i \equiv b_i - c_i^T$. Then for any pair of bidders i and j, we have:

$$\begin{aligned} c_i^B - c_j^B &= (c_i^T - c_j^T) - (c_i^A - c_j^A) \\ &= (b_i - b_j) - (m_i - m_j) - (c_i^A - c_j^A) \end{aligned}$$

Now let i be the winning bidder W and let j be the bidder with the lowest base cost:

$$\begin{aligned} \max_j (c_W^B - c_j^B) &= \max_j \big((b_W - b_j) - (m_W - m_j) - (c_W^A - c_j^A)\big) \\ &\leq \max_j (b_W - b_j) - \min_j (m_W - m_j) - \min_j (c_W^A - c_j^A) \\ &\leq \max_j (b_W - b_j) - \min_j (c_W^A - c_j^A) \end{aligned}$$

where the last line follows from either constant markups, or from markups decreasing in score (so the winning bidder always has the highest markup). Both of the expressions on the RHS can be simulated contract-by-contract, and then averaged to get an upper bound on the average increase in base costs.

4.3 Robustness and Sensitivity Checks

Last, we present a series of robustness tests on the counterfactual results. In tables 5, 6, 7, 8, and 9, we reproduce the main counterfactual results under alternative assumptions. In table 5, we re-estimate the structural model without any controls derived from previous regressions (i.e. the participation residual in the bidding regression, and the participation and bid residuals in the probability of winning regression). This would be appropriate if there was no selection on unobservables: for example, conditional on participation all bidders drew independent types from the same distribution. In table 6 we shut down all forms of observable asymmetry between the bidders (e.g. distance from job site, size of firm), and basically assume bidder symmetry throughout. What we see is that the results are very similar to those in the main paper. The general trend is that the welfare gains from A+B

assignment are somewhat muted, presumably because this shuts down one channel for these policies to be effective: by matching firms that are typically better at contract acceleration (close firms, big firms) with A+B contracts.

The next set of robustness checks explores the sensitivity of the estimates to large changes in the assumptions. In Table 7 we re-run the simulations assuming that the daily welfare costs to commuters were only a third of our estimates. This makes them more comparable to the usercosts assigned by Caltrans. Predictably, the welfare gains are far smaller. In particular, if our welfare estimates were far too high, then the policy of only assigning A+B status to the contracts that inflict the greatest negative externality on commuters (third column) doesn't perform that much worse than the budget policy (fifth column). Still, the broad conclusions are unchanged: expanding the program to encompass all contracts would be helpful, even if usercosts were set at 10% of the externality.

In Tables 8 and 9 we change the key parameter in the structural estimation α, which governs the convexity of the marginal cost of acceleration curve. In Table 8 we double it, making acceleration costs way more convex, whereas in Table 9 we halve it, making them flatter. Having forced a particular value for α, we estimate the bidding regression as before with this constraint, and then proceed as in the main paper. The big difference is in the relative effectiveness of the "budget" policy. When costs are more convex, the budget policy is exceptionally effective, obtaining 90% of the welfare gain of the efficient policy at only 13.5% of the acceleration cost to contractor. However when they are flatter, the budget policy doesn't do quite as well, getting 65% welfare for 10.4% of the cost. This illustrates the importance of getting a precise estimate of the convexity of the cost curve.

Table 5: Counterfactual Welfare Estimates: No Selection Controls

A+B Contracts	Current	Current	Welfare > $100K	All	All
Usercost	Current	100% Welfare	100% Welfare	100% Welfare	10% Welfare
Days Bid / Engdays (%)	69.03	63.05	52.42	62.55	74.20
	(65.47,72.37)	(55.53,69.09)	(44.02,67.03)	(54.12,73.29)	(62.14,85.25)
Mean Commuter Gain ($M)	3.98	5.35	9.20	2.24	1.60
	(3.55,4.42)	(3.95,7.14)	(6.62,11.41)	(1.73,2.81)	(1.13,1.97)
Mean Acc. Cost ($M)	0.380	0.927	0.710	0.304	0.037
	(0.054,0.788)	(0.120,1.837)	(0.088,1.485)	(0.039,0.615)	(0.006,0.063)
Change in Base Cost ($M)	0.231	0.575	0.425	0.179	0.025
	(0.130,0.336)	(0.208,0.929)	(0.149,0.747)	(0.067,0.288)	(0.014,0.031)
Mean Net Gain ($M)	3.37	3.85	8.07	1.76	1.53
	(2.84,3.91)	(3.30,4.56)	(5.92,9.65)	(1.42,2.12)	(1.04,1.90)
Total Cost Increase ($M)	47.01	115.66	94.21	320.07	40.93
	(14.61,80.97)	(31.37,208.28)	(23.89,167.98)	(84.76,567.35)	(12.57,59.07)
Total Net Gain ($M)	259.52	296.59	669.70	1163.43	1015.47
	(218.50,301.24)	(253.78,350.77)	(491.21,800.95)	(942.73,1401.37)	(689.50,1258.65)

Counterfactual welfare results under different policies, estimated without any of controls for selection (i.e. "participation residual", "bid residual"). The first column is simulated outcomes under the observed A+B policy, averaged across all A+B contracts. In the second column, we maintain the observed A+B assignment but the usercosts are set equal to the estimated commuter welfare; in the third, all contracts with commuter welfare above $100 000 per day are assigned A+B status and usercost is equal to the welfare; in column four all contracts are A+B, with usercost equal to welfare; and in the last column, all are A+B, but the usercost is only 10% of the welfare. "Commuter Gain" for any contract is calculated as the product of commuter welfare and the difference between the days bid and 92.5% of the engineer's days, since A+B contracts are completed on time and standard contracts are typically completed 7.5% early. "Acceleration Cost" is the estimated additional cost to the winning contractor of this accelerated construction schedule. "Change in Base Cost" is an upper bound on the change in the base cost of completing the project by selecting the highest scoring rather the bidder with the lowest base cost. "Net Gain" is a lower bound on the welfare gain from A+B assignment. "Cost Increase" is the sum of acceleration costs and the change in base cost. In all cases, mean results are averaged across simulations and A+B contracts. Totals are obtained by summing the relevant statistics across all A+B contracts. 95% confidence intervals are given in parentheses, and are generated by bootstrapping the bid and participation regressions in the main text, and taking the 2.5th and 97.5th percentiles of the simulated results based on the bootstrapped coefficients.

Table 6: Counterfactual Welfare Estimates: No Observable Bidder Asymmetry

A+B Contracts	Current	Current	Welfare > $100K	All	All
Usercost	Current	100% Welfare	100% Welfare	100% Welfare	10% Welfare
Days Bid / Engdays (%)	68.94	63.20	53.87	62.05	73.59
	(65.39,72.41)	(55.45,69.40)	(44.97,67.10)	(53.74,72.92)	(63.82,85.09)
Mean Commuter Gain ($M)	3.95	5.28	8.82	2.18	1.56
	(3.49,4.42)	(3.88,7.14)	(6.37,10.94)	(1.66,2.74)	(1.09,1.95)
Mean Acc. Cost ($M)	0.368	0.905	0.689	0.295	0.036
	(0.082,0.786)	(0.193,1.909)	(0.148,1.443)	(0.060,0.628)	(0.009,0.061)
Change in Base Cost ($M)	0.253	0.557	0.383	0.171	0.024
	(0.145,0.347)	(0.241,0.885)	(0.166,0.652)	(0.078,0.284)	(0.015,0.031)
Mean Net Gain ($M)	3.33	3.81	7.75	1.72	1.50
	(2.68,3.90)	(3.30,4.57)	(5.72,9.53)	(1.37,2.08)	(1.01,1.90)
Total Cost Increase ($M)	47.83	112.63	88.99	308.16	39.18
	(18.54,82.20)	(40.97,205.34)	(31.29,160.68)	(110.86,557.44)	(15.67,57.13)
Total Net Gain ($M)	256.46	293.70	643.44	1136.47	993.40
	(206.59,300.07)	(254.33,351.99)	(474.69,791.13)	(906.83,1378.51)	(667.69,1256.05)

Counterfactual welfare results under different policies, estimated assuming bidder symmetry up to unobserved type (i.e. no observable controls for capacity, distance from job site etc). The first column is simulated outcomes under the observed A+B policy, averaged across all A+B contracts. In the second column, we maintain the observed A+B assignment but the usercosts are set equal to the estimated commuter welfare; in the third, all contracts with commuter welfare above $100 000 per day are assigned A+B status and usercost is equal to the welfare; in column four all contracts are A+B, with usercost equal to welfare; and in the last column, all are A+B, but the usercost is only 10% of the welfare. "Commuter Gain" for any contract is calculated as the product of commuter welfare and the difference between the days bid and 92.5% of the engineer's days, since A+B contracts are completed on time and standard contracts are typically completed 7.5% early. "Acceleration Cost" is the estimated additional cost to the winning contractor of this accelerated construction schedule. "Change in Base Cost" is an upper bound on the change in the base cost of completing the project by selecting the highest scoring rather the bidder with the lowest base cost. "Net Gain" is a lower bound on the welfare gain from A+B assignment. "Cost Increase" is the sum of acceleration costs and the change in base cost. In all cases, mean results are averaged across simulations and A+B contracts. Totals are obtained by summing the relevant statistics across all A+B contracts. 95% confidence intervals are given in parentheses, and are generated by bootstrapping the bid and participation regressions in the main text, and taking the 2.5th and 97.5th percentiles of the simulated results based on the bootstrapped coefficients.

Table 7: Counterfactual Welfare Estimates: Sensitivity to Welfare Estimates

A+B Contracts	Current	Current	Welfare > $100K	All	All
Usercost	Current	100% Welfare	100% Welfare	100% Welfare	10% Welfare
Days Bid / Engdays (%)	68.81	68.92	52.87	67.45	79.23
	(65.54,72.67)	(65.86,71.94)	(45.29,67.58)	(58.68,76.96)	(63.34,90.88)
Mean Commuter Gain ($M)	1.34	1.53	2.93	0.67	0.42
	(1.17,1.50)	(1.26,1.80)	(2.04,3.52)	(0.50,0.79)	(0.18,0.61)
Mean Acc. Cost ($M)	0.383	0.385	0.405	0.128	0.012
	(0.052,0.781)	(0.050,0.795)	(0.050,0.806)	(0.017,0.250)	(0.002,0.015)
Change in Base Cost ($M)	0.253	0.292	0.338	0.098	0.007
	(0.152,0.350)	(0.166,0.390)	(0.188,0.442)	(0.059,0.128)	(0.004,0.008)
Mean Net Gain ($M)	0.70	0.85	2.19	0.44	0.40
	(0.24,1.14)	(0.60,1.16)	(1.41,2.82)	(0.29,0.58)	(0.17,0.60)
Total Cost Increase ($M)	49.02	52.11	61.71	149.67	12.23
	(16.39,81.98)	(19.10,89.87)	(22.83,100.03)	(54.46,240.46)	(3.64,15.21)
Total Net Gain ($M)	54.04	65.51	181.36	292.52	264.03
	(18.43,87.58)	(46.39,89.04)	(116.74,233.86)	(189.22,386.70)	(110.24,398.92)

Counterfactual welfare results under different policies, obtained after dividing our welfare estimates by 3. The first column is simulated outcomes under the observed A+B policy, averaged across all A+B contracts. In the second column, we maintain the observed A+B assignment but the usercosts are set equal to the estimated commuter welfare; in the third, all contracts with commuter welfare above $100 000 per day are assigned A+B status and usercost is equal to the welfare; in column four all contracts are A+B, with usercost equal to welfare; and in the last column, all are A+B, but the usercost is only 10% of the welfare. "Commuter Gain" for any contract is calculated as the product of commuter welfare and the difference between the days bid and 92.5% of the engineer's days, since A+B contracts are completed on time and standard contracts are typically completed 7.5% early. "Acceleration Cost" is the estimated additional cost to the winning contractor of this accelerated construction schedule. "Change in Base Cost" is an upper bound on the change in the base cost of completing the project by selecting the highest scoring rather the bidder with the lowest base cost. "Net Gain" is a lower bound on the welfare gain from A+B assignment. "Cost Increase" is the sum of acceleration costs and the change in base cost. In all cases, mean results are averaged across simulations and A+B contracts. Totals are obtained by summing the relevant statistics across all A+B contracts. 95% confidence intervals are given in parentheses, and are generated by bootstrapping the bid and participation regressions in the main text, and taking the 2.5th and 97.5th percentiles of the simulated results based on the bootstrapped coefficients.

Table 8: Counterfactual Welfare Estimates: Less convex acceleration costs

A+B Contracts	Current	Current	Welfare > $100K	All	All
Usercost	Current	100% Welfare	100% Welfare	100% Welfare	10% Welfare
Days Bid / Engdays (%)	68.83	56.64	47.58	61.19	80.24
	(65.62,72.72)	(53.79,60.91)	(39.47,65.58)	(55.32,72.08)	(74.86,86.53)
Mean Commuter Gain ($M)	4.00	6.72	10.55	2.63	1.42
	(3.55, 4.50)	(6.15, 7.17)	(8.23,12.06)	(2.25, 2.92)	(1.05, 1.68)
Mean Acc. Cost ($M)	0.676	1.601	1.234	0.540	0.059
	(0.595,0.721)	(1.441,1.680)	(0.986,1.290)	(0.477,0.559)	(0.052,0.065)
Change in Base Cost ($M)	0.301	0.973	0.770	0.298	0.027
	(0.226,0.362)	(0.753,1.223)	(0.530,1.054)	(0.213,0.381)	(0.022,0.031)
Mean Net Gain ($M)	3.03	4.15	8.54	1.80	1.33
	(2.58, 3.46)	(3.49, 4.63)	(6.21,10.09)	(1.41, 2.07)	(0.97, 1.59)
Total Cost Increase ($M)	75.25	198.25	166.31	554.48	57.07
	(66.12,81.10)	(175.97, 214.03)	(141.09, 180.19)	(478.18, 593.78)	(50.10,62.61)
Total Net Gain ($M)	233.13	319.24	709.12	1189.55	880.59
	(198.96, 266.26)	(268.36, 356.48)	(515.22, 837.35)	(933.77,1371.74)	(641.88,1053.94)

Counterfactual welfare results under different policies. As a sensitivity check, this imposes $\alpha = 1.81$ in the structural estimation (half our original estimate), implying less convex marginal costs of acceleration. The first column is simulated outcomes under the observed A+B policy, averaged across all A+B contracts. In the second column, we maintain the observed A+B assignment but the usercosts are set equal to the estimated commuter welfare; in the third, all contracts with commuter welfare above $100 000 per day are assigned A+B status and usercost is equal to the welfare; in column four all contracts are A+B, with usercost equal to welfare; and in the last column, all are A+B, but the usercost is only 10% of the welfare. "Commuter Gain" for any contract is calculated as the product of commuter welfare and the difference between the days bid and 92.5% of the engineer's days, since A+B contracts are completed on time and standard contracts are typically completed 7.5% early. "Acceleration Cost" is the estimated additional cost to the winning contractor of this accelerated construction schedule. "Change in Base Cost" is an upper bound on the change in the base cost of completing the project by selecting the highest scoring rather the bidder with the lowest base cost. "Net Gain" is a lower bound on the welfare gain from A+B assignment. "Cost Increase" is the sum of acceleration costs and the change in base cost. In all cases, mean results are averaged across simulations and A+B contracts. Totals are obtained by summing the relevant statistics across all A+B contracts. 95% confidence intervals are given in parentheses, and are generated by bootstrapping the bid and participation regressions in the main text, and taking the 2.5th and 97.5th percentiles of the simulated results based on the bootstrapped coefficients.

Table 9: Counterfactual Welfare Estimates: More convex acceleration costs

A+B Contracts	Current	Current	Welfare > $100K	All	All
Usercost	Current	100% Welfare	100% Welfare	100% Welfare	10% Welfare
Days Bid / Engdays (%)	68.74	65.87	54.06	62.02	68.55
	(65.69,72.33)	(62.91,69.45)	(46.37,68.91)	(54.64,73.31)	(61.63,78.21)
Mean Commuter Gain ($M)	4.03	4.70	8.64	2.07	1.74
	(3.46,4.47)	(4.09,5.18)	(6.06,10.17)	(1.59,2.39)	(1.29,2.07)
Mean Acc. Cost ($M)	0.201	0.465	0.356	0.153	0.021
	(0.176,0.213)	(0.414,0.489)	(0.289,0.377)	(0.128,0.157)	(0.018,0.022)
Change in Base Cost ($M)	0.214	0.648	0.550	0.209	0.021
	(0.154,0.264)	(0.333,0.891)	(0.271,0.767)	(0.115,0.280)	(0.016,0.024)
Mean Net Gain ($M)	3.62	3.59	7.74	1.71	1.70
	(3.03,4.06)	(2.81,4.13)	(5.06,9.46)	(1.24,2.05)	(1.25,2.03)
Total Cost Increase ($M)	31.98	85.69	75.18	239.19	28.17
	(26.66,35.65)	(61.50,105.64)	(49.77,91.68)	(170.21,276.57)	(23.56,29.95)
Total Net Gain ($M)	278.38	276.57	642.31	1133.96	1126.67
	(233.17,312.35)	(216.59,318.05)	(420.28,785.09)	(819.70,1359.63)	(825.02,1340.88)

Counterfactual welfare results under different policies. As a sensitivity check, this imposes $\alpha = 7.27$ in the structural estimation (twice our original estimate), implying more convex marginal costs of acceleration. The first column is simulated outcomes under the observed A+B policy, averaged across all A+B contracts. In the second column, we maintain the observed A+B assignment but the usercosts are set equal to the estimated commuter welfare; in the third, all contracts with commuter welfare above $100 000 per day are assigned A+B status and usercost is equal to the welfare; in column four all contracts are A+B, with usercost equal to welfare; and in the last column, all are A+B, but the usercost is only 10% of the welfare. "Commuter Gain" for any contract is calculated as the product of commuter welfare and the difference between the days bid and 92.5% of the engineer's days, since A+B contracts are completed on time and standard contracts are typically completed 7.5% early. "Acceleration Cost" is the estimated additional cost to the winning contractor of this accelerated construction schedule. "Change in Base Cost" is an upper bound on the change in the base cost of completing the project by selecting the highest scoring rather the bidder with the lowest base cost. "Net Gain" is a lower bound on the welfare gain from A+B assignment. "Cost Increase" is the sum of acceleration costs and the change in base cost. In all cases, mean results are averaged across simulations and A+B contracts. Totals are obtained by summing the relevant statistics across all A+B contracts. 95% confidence intervals are given in parentheses, and are generated by bootstrapping the bid and participation regressions in the main text, and taking the 2.5th and 97.5th percentiles of the simulated results based on the bootstrapped coefficients.

References

Abadie, Alberto, David Drukker, Jane Herr, and Guido Imbens, "Implementing Matching Estimators for Average Treatment Effects in Stata," *The Stata Journal*, 2004, *4* (3), 290–311.

Imbens, Guido and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 2008, *142*, 615–635.