Highway Procurement and the Stimulus Package: Identification and Estimation of Dynamic Auctions with Unobserved Heterogeneity

Jorge Balat*
Johns Hopkins University

November 27, 2013

Abstract

In the highway procurement market, if firms' marginal costs are intertemporally linked, the pace at which the government releases new projects over time will have an effect on the prices it pays. This paper investigates the effects of the American Recovery and Reinvestment Act on equilibrium prices paid by the government for highway construction projects using data from California. I develop a structural dynamic auction model that allows for intertemporal links in firms' marginal costs, project level unobserved heterogeneity, and endogenous participation. I show that the model is nonparametrically identified combining ideas from the control function and measurement error literatures. I find that the accelerated pace of the Recovery Act projects imposed a sizable toll on procurement prices, especially on the procurement cost of projects not funded by the stimulus money.

1 Introduction

The American Recovery and Reinvestment Act (ARRA) of 2009 stipulated a large injection of funds (over $800 billion) into the economy in a short period of time.[1] This

*I am indebted to Phil Haile and Steve Berry for support and advice. I thank John Asker, Ariel Pakes, Martin Pesendorfer, Yingyao Hu, Elena Krasnokutskaya, Taisuke Otsu, Jean Tirole, and seminar participants at BU, Bocconi, FGV-São Paulo, Harvard, LSE, UPenn, Princeton, Pompeu Fabra, PUC-Rio, Toulouse, Wharton, and Yale for helpful comments. I also thank Eduardo Souza-Rodriguez, Ted Rosenbaum, Myrto Kalouptsidi, Sukjin Han, Priyanka Anand for extensive discussions, and Earl Seaberg and Rich Stone at Caltrans for clarifying conversations about the data. All errors and opinions remain my own.

[1]Throughout the paper, I will refer to the American Recovery and Reinvestment Act, as the stimulus package.

paper investigates the effects of this demand expansion on equilibrium prices paid by the government for highway construction projects. Using data from California, I answer the following questions: (1) How much were the costs of these projects driven up by the accelerated pace of new projects? (2) What was the effect of the demand expansion on the prices of other state projects that came afterwards? and (3) What was the effect on efficiency? The first question aims at quantifying the trade-off the government faces between getting money in people's hands right away and getting more public goods out of the stimulus funds. The second question makes explicit certain overlooked costs associated with the stimulus funds received by the states. The third question investigates the effect of the stimulus package on the cost of resources to society.

Road construction projects are awarded by auctions, therefore, I develop a structural dynamic auction model in order to answer the questions raised above. The model builds on three key features. First, I allow for an inter-temporal link in firm's marginal costs, i.e., firms' current costs can be affected by their committed resources from uncompleted projects awarded in previous periods. Second, I allow for auction-level unobserved heterogeneity, a project-level factor that affects firms' behavior but is not observed by the econometrician. Finally, I allow for endogenous firm participation, i.e., I let firms' participation decisions depend on project characteristics (both observed and unobserved). I apply this model to highway procurement data from California to estimate the structural parameters, and then perform simulations to assess the issues raised above.

The basic economic ideas behind my questions are simple. When firms have upward sloping marginal costs[2] and projects take more than one period to complete, there will be a dynamic linkage between projects awarded in previous periods and the price of projects in the current period. In other words, firms' backlogs of uncompleted projects will affect the current period supply curve. There are many channels that may affect the supply. Firms operating with a high backlog will likely bid less aggressively for new projects, since taking an additional contract may require, for example, renting equipment, which in turn increases the firm's total costs. Moreover, even firms with low backlogs may end up bidding less aggressively as well, through a strategic effect, if rivals are constrained.[3] There is also a competition effect, since high backlogs may

[2]It could also be the case that firms face downward sloping marginal cost curves if, for example, they experience learning-by-doing. In this paper, I leave firms' cost structure unrestricted and let the data tell me how marginal costs and backlogs are related. I do find that marginal costs are upward sloping.

[3]Throughout the paper I may refer to a firm being constrained or unconstrained to imply that the firm has a high or low backlog, respectively.

deter firms from participating in the auctions, thus reducing the number of bidders and lessening competition. As a result, for any given project, the higher the backlogs, the higher the price paid by the government.

The questions raised in this paper are important because the government had two objectives behind the stimulus package. On the one hand, it wanted to jumpstart the economy as soon as possible, and on the other, it wanted to invest in transportation infrastructure. But if the supply is slow to react, there is a trade-off between the two objectives. This can be explained in the context of my model in the following way. Besides choosing the size of the expansion, the government can choose when and how to pace the release of new projects. If it delays new projects, backlogs decrease as the previously awarded ongoing projects progress. Therefore, we expect lower prices for new projects. Given a fixed amount of money to spend, this translates into more public goods acquired. The government then has to decide whether to spend the money right away at the cost of fewer public goods, or to postpone (or slow down) the expansion and get more public goods.[4]

There is also another effect of the demand expansion that affects state governments. Conditional on the current backlog, the projects funded by the stimulus money raise the backlog level in future periods, hence future projects funded by the state will face higher prices. This is, in fact, a *hidden* cost of the stimulus funds received by the states.

Finally, from a welfare perspective, if one were to ignore the cost of raising the money to finance the stimulus package, one should only care about the cost of the resources used to build or repair the roads. Due to the effect of backlogs on firms' costs, we should expect an inefficiency generated by the stimulus funds.

The highway construction sector and the stimulus money targeted at it are economically important. In particular, highway construction procurement projects auctioned off by the government accounted for \$66 billion in 2007,[5] and \$50 billion of the stimulus money was targeted at transportation infrastructure (out of which, \$30 billion was allocated to construction and repair of highways, roads and bridges, the biggest single line infrastructure item in the final bill).

I turn now to a description of the methodology used in this paper. Initially, the empirical literature on auctions assumed away unobserved heterogeneity because

[4]A related policy issue concerns the optimal project timing to minimize procurement costs conditional on achieving some stimulus objective. This analysis is beyond the scope of the present paper and is left for future research.

[5]Data from the US Census Bureau. The figure comprises highway, street and bridges projects contracted by Federal, State and local governments.

identification failed otherwise. But this assumption is likely to be violated in many settings and recent work has paid increasingly attention to it. My estimation approach uses concepts from the control function and measurement error literatures to show that the structural model is nonparametrically identified under the presence of unobserved heterogeneity. I use the first order condition at the bidding stage to express each firm's private cost as a function of its bid, the conditional distribution of equilibrium bids,[6] and the value function representing the discounted sum of future payoffs. The main challenge is to show the identification of the distribution function of equilibrium bids conditional on the unobserved heterogeneity, and of the distribution of unobserved heterogeneity.

My estimation approach combines several key ideas. First, provided that the distribution of unobserved heterogeneity is identified, I can write the value function as a function of the distribution of equilibrium bids, as proposed by Jofre-Bonet and Pesendorfer (2003) (JP hereafter) for a dynamic model without unobserved heterogeneity. A second key idea is similar in spirit to the control function approach (see Chesher (2003) and Imbens and Newey (2009)). The classic approach relies on an equation that relates an endogenous observed outcome to the unobserved factor. With a strict monotonicity assumption, the relationship can then be inverted and used to control for the unobserved factor directly. I depart from this method by allowing the control function to be an "imperfect" one. That is, I do not require the observables in the relationship to be a sufficient statistic for the unobserved heterogeneity.[7] Rather, I exploit features of the procurement setting that provide a second "imperfect" control function. The information obtained from these two noisy controls then resembles a measurement error problem where we have access to multiple measurements. I attain identification of the conditional distribution functions using the results in Hu (2008) for nonlinear models with misclassification error.

This paper contributes to the auction literature in several ways. I improve on the method of JP by controlling for unobserved heterogeneity, and by relaxing the assumptions on firms' participation decisions allowing for endogenous participation. To my knowledge, this is the first attempt to control for unobserved heterogeneity in a dynamic auction model. I also relax the structural assumptions in the control function approaches used previously in the auction literature (see Haile, Hong, and Shum

[6]Conditional on observed and unobserved auction level factors, and firms' state variables. To economize on terminology from now on, when I say distribution of equilibrium bids or distribution of private values, I always refer to the conditional distribution.

[7]There are various reasons why this could be the case. For example, the agent's decision outcome that we are exploiting may be based on a noisy version of the unobserved factor that we need to control for, or other unobservable shocks may also affect the outcome.

(2006) —HHS hereafter— and Roberts (2011)) by allowing the control function to be an imprecise measure of the unobserved heterogeneity. In a recent attempt to control for unobserved factors in a static context, Krasnokutskaya (2011) uses a deconvolution method from the measurement error literature to identify an independent private values model, in which unobserved heterogeneity enters linearly in the valuation and is independent of the idiosyncratic components of bidders' values. In my model, I do not require specifying the functional form for the relationship between the valuation and the unobserved factor, and I let the idiosyncratic component of the firm's valuation to be correlated with the unobserved component.[8][9]

The estimation results suggest that both the effects of backlog and unobserved heterogeneity are important. I find that increasing the unobserved heterogeneity from its lowest level to its highest level rises the cost of a firm by 51%, and the equilibrium winning bid by 15%. Furthermore, although monotone, the effects are nonlinear in the level of unobserved heterogeneity. Changing the backlog level of one regular firm from a low level to a high one, increases the cost of that firm by 7.4%, and the equilibrium price in the auction increases by 4.7%.

From a policy perspective, this paper contributes to the discussion about the stimulus package by raising questions that have not been addressed yet. I quantify costs associated with the stimulus projects that may help in future policy making. Using the estimates of the structural parameters of the model in counterfactual simulations indicate that the government paid prices for stimulus funded projects that are 6.2% higher (and prices for other projects that are 4.8% higher) due to the effect of the stimulus projects on firms' backlogs. These results imply that the government could have acquired $335 million worth of extra road projects (or 19.7% of the stimulus money received by California) by forgoing any stimulus effect from ARRA. By no means does this suggest that the $335 million were "lost," since the money was actually transferred to the firms, thus serving the government's stimulus objective. The result just makes explicit that part of the demand expansion went into higher prices rather than quantities. Moreover, I also show that even though California presents features that may distinguish it from other states, the results are robust and can be extrapolated to other applications.

[8]Although in this paper I consider a dynamic independent private values model, the identification result holds without modification in a static affiliated private values context. Thus, this is also an extension of Krasnokutskaya's method. I am currently working on an extension of the dynamic model with affiliated private values.

[9]Even though I appeal to additional structure to get two noisy measures of the control function, it can be substituted for two discretized bids (see An, Hu, and Shum (2010)). Therefore, my method does not necessarily entail a greater burden on data or structure than other methods.

In a separate set of simulations I find that the benefit (in terms of lower prices) of delaying the stimulus projects by 3 months reaches \$44 million (or 2.6% of the stimulus funds). If the government delays the projects by 6 months instead, the benefit totals \$62 million (or 3.7% of the stimulus funds). Finally, results regarding the inefficiency created by the stimulus projects show that the cost of the resources used in projects funded by the stimulus money and by other sources increased by 2.8% on average.

The rest of the paper proceeds as follows. I begin with a literature review in Section 2. Section 3 summarizes features of the stimulus package and its potential effects. Section 4 introduces the data and details of the procurement process. The structural model is presented in Section 5. Identification and estimation are discussed in Sections 6 and 7, respectively. Section 8 presents the estimation results, and Section 9 the simulation results. Finally, Section 10 concludes.

2 Literature Review

This paper is related to several strands in the auction literature. In this section I summarize some of the recent findings.

Dynamic models. The only papers I am aware of that estimate a dynamic auction model are JP and Groeger (2010). JP look at how capacity constraints affect bidding behavior. Although my model is based on JP, I extend it to allow for endogenous participation and unobserved heterogeneity. Groeger (2010) estimates a dynamic model with endogenous participation. While his interest is in the dynamic synergies that result from participation I focus on the dynamic effect of backlogs.

Nonparametric estimation of private values models. The seminal work of Guerre, Perrigne, and Vuong (2000) studies the identification and estimation of a first-price symmetric independent private values (IPV) model. Li, Perrigne, and Vuong (2002) extend the result to the affiliated private values (APV) setting and Li, Perrigne, and Vuong (2000) to the conditionally independent private values. These papers assume away the existence of unobserved heterogeneity and rely on an invertible mapping between the distribution of bidders values and the distribution of observed bids.

Asymmetric bidders. More closely related to my setting is the literature that estimates structural first-price auction models with asymmetric bidders. This strand includes Bajari (1997), Bajari (2001), Hong and Shum (2002), Bajari and Ye (2003), Campo, Perrigne, and Vuong (2003), JP, Flambard and Perrigne (2006), Einav and Esponda (2008), and Athey, Levin, and Seira (2011).

Unobserved heterogeneity. Also closely related is the increasing literature on nonparametric identification of auction models with unobserved heterogeneity. Campo, Perrigne, and Vuong (2003), HHS, and Guerre, Perrigne, and Vuong (2009) use information on the number of bidders. The first paper assumes that the number of bidders is a sufficient statistic for the unobserved auction heterogeneity. HHS use a control function approach to account for the unobserved heterogeneity. They assume endogenous participation and model the number of bidders as a strictly increasing function of the unobserved heterogeneity. Roberts (2011) also uses a control function approach. He assumes that reserve prices are monotonic in the unobserved heterogeneity. Guerre, Perrigne, and Vuong (2009) build on the methodology of HHS to identify an IPV model with risk averse bidders and unobserved heterogeneity based on exclusion restrictions derived from bidders endogenous participation. Hong and Shum (2002) and Athey, Levin, and Seira (2011) take a parametric approach. The first paper assumes that the median of the bid distribution is distributed Normal with mean and variance dependent on the number of bidders, while the second one assumes a parametric form for the distribution of bids conditional on the unobserved heterogeneity and that the unobserved heterogeneity follows a Gamma distribution.

Other approaches use results from the measurement error literature. Krasnokutskaya (2011) uses a deconvolution method to identify an IPV model in which unobserved heterogeneity enters linearly in the bidder's valuation and is independent of the idiosyncratic components of bidders' values. More recently, An, Hu, and Shum (2010) and Hu, McAdams, and Shum (2011) rely on new results in the econometric literature on nonclassical measurement error by Hu (2008) and Hu and Schennach (2008). They extend the results in Krasnokutskaya (2011) by allowing the unobserved heterogeneity to be nonseparable from bidders' valuations. Krasnokutskaya (2011), An, Hu, and Shum (2010) and Hu, McAdams, and Shum (2011) rely on the IPV setting and require data on multiple bids from the same auction to identify the distribution of bidders' valuations. The three papers consider a static auction model and none of them deals with endogenous participation.

My identification approach combines features of both the control function techniques and of the recent papers that exploit the nonclassical measurement error results and I apply it to a dynamic model with endogenous participation. By combining the two methods I am able to extend the results from both approaches. First, I can relax the structural assumptions in the control function approach by allowing the control function to be an imprecise measure of the unobserved heterogeneity. Second, I can extend the results in Hu, McAdams, and Shum (2011) to the APV setting (although in the present paper I focus on a dynamic model with IPV, in ongoing work I show nonparametric identification of a static model with APV). Furthermore, within the static IPV paradigm although my method requires data on two additional endogenous outcomes it only requires having data on the winning bid in an auction.

Highway procurement. Finally, my paper is related to earlier work on the analysis of highway contracts. For example, Porter and Zona (1993) and Bajari and Ye (2003) are interested in detecting collusion among bidders, Hong and Shum (2002) assess the winner's curse, JP find evidence of capacity constraints, Bajari and Tadelis (2001) and Bajari, Houghton, and Tadelis (2006) look at the implications of incomplete procurement contracts, Krasnokutskaya (2011) finds evidence of unobserved heterogeneity, Krasnokutskaya and Seim (2011) study bid preference programs and participation, Li and Zheng (2009) and Einav and Esponda (2008) look at endogenous participation and its effect on procurement cost.

3 Stimulus Package and Its Effects

The American Recovery and Reinvestment Act is a job and economic stimulus bill intended to help states and the nation restart their economies and stimulate employment after the worst economic downturn in over 70 years. In drafting this bill, Congress recognizes that investment in transportation infrastructure is one of the best ways to create and sustain jobs, stimulate economic development, and leave a legacy to support the financial well-being of the generations to come. The intent and language of the bill responded to the urgency of the economic situation by tasking state departments of transportation and other transportation stakeholders to quickly move forward with transportation infrastructure projects.

Nationally, the bill provided $50 billion for transportation infrastructure, out of which $30 billion was allocated to construction and repair of highways, roads and bridges. The latter constituted the biggest single line infrastructure item in the final bill. The state of California received the most funds of any state, with approximately $2.6 billion awarded for highways, local streets and roads projects,[10] and $1.07 billion for transit projects in addition to other discretionary funds.[11] The stimulus funds became the primary focus of California's Department of Transportation (Caltrans),

[10]About $900 million correspond to local streets and the funds were administered by the local governments. I exclude these projects from my analysis.

[11]In January 2010, California was awarded more than $2.3 billion for its high-speed intercity rail, the largest allocation in the nation. The state also received an additional $130 million in new funding for four highway, local street, rail and port projects across the state from the Recovery Acts Transportation Investment Generating Economic Recovery (TIGER) Grant program.

and in March 2009, California modified existing law to allow projects to start sooner. Even though under the guidelines of the Recovery Act states were given 120 days to obligate half of their federal stimulus transportation funding to projects, California obligated half the funds two months ahead of the deadline. Furthermore, California was the first state in the nation to obligate $1 billion (by May 2009) and $2 billion (by September 2009). The funding, was fully obligated on February 18, 2010, and was designated to 907 projects (516 projects worth $2.5 billion were already awarded to begin work by then).

In summary, the government had two objectives behind the stimulus package which are explicit in the name of the Act. On the one hand, it aims at jumpstarting the economy as soon as possible, and on the other, it aims at investing in transportation infrastructure. But when the supply is slow to react, there is a trade-off between the two objectives. To some extent one can reinterpret the result in Goolsbee (1998) in terms of this trade-off. Using data from R&D government expenditure, he finds that the majority of the expansion in R&D expenditure goes directly into higher wages, an increase in the price rather than the quantity of inventive activity. This is due to an inelastic supply of scientific and engineering talent. Therefore, even though the money from the expansion still goes into the people's hands —scientists in his case— it does not necessarily translate into more public goods.

In the highway procurement setting, the trade-off is caused by the inter-temporal linkage of firms' costs (through their backlogs) and upward sloping marginal cost curves. JP documents the fact that previously won and uncompleted contracts affect a firm's current costs. Since the duration of highway paving contracts is typically several months, winning a large contract may commit some of the bidder's machines and resources for the duration of the contract. Although a firm can pay overtime wages, hire additional workers and rent additional equipment, this may increase total cost. Therefore the cost of taking on an additional contract is increasing in the firm's backlog.[12] As a result, for any given project, the higher the current backlogs are the higher the price paid by the government.

In addition to the *direct* effect of backlogs on prices, there are other *indirect* channels which affect the equilibrium price of a project. Through a strategic effect, even firms currently operating with a low level of backlog may end up bidding less aggressively if rivals are constrained. In a static context, Maskin and Riley (2000) and Cantillon (2008) find that asymmetries in bidders' costs soften competition since

[12] In a static context, Bajari and Ye (2003), DeSilva, Dunne, and Kosmopoulou (2003), and Bajari, Houghton, and Tadelis (2006) also find that firms bid less aggressively as backlog increases.

low cost firms react to high cost firms participating in the auction by strategically increasing their bids. The asymmetries in my model arise from firms having different levels of backlogs. Furthermore, there is a competition effect affecting prices. High backlogs may deter firms from participating in the auctions, thus reducing the number of bidders. A key feature of my model, as opposed to the one in JP, is that it considers endogenous entry, thus taking the latter effect into account.

From the stimulus objective perspective, an increase in the prices of projects poses no problem since money still reaches the firms. However, given a fixed budget, higher prices clash with the investment objective since the government can buy fewer public goods. Moreover, from a welfare point of view, if we were to ignore the cost of raising the money to finance the stimulus package (i.e., suppose there is no surplus lost when the government raises taxes), we should only care about the cost of the resources used to build or repair the roads. Since higher backlogs increase firms' costs, the demand expansion generates an inefficiency.

In the highway procurement setting, the government does not only choose the amount of the demand expansion, but has another choice variable at hand: when and how to pace the release of new projects. Pacing matters because the current level of firms' backlogs is a function of previously awarded contracts. To illustrate the issue, think of an extreme case. Suppose the government does not announce any new projects. As previously awarded projects progress, the backlog levels decay naturally over time. As was discussed above, the lower the backlog, the lower the prices for new projects. The longer the government waits to release the stimulus funded projects, the lower its prices, and given a fixed amount to spend, the higher the quantities of public goods (i.e., more roads) it can purchase. But on the other hand, firms may receive the funds too late, defeating the government's first objective. In the words of Caltrans Director Will Kempton, it appears that the priority was putting the money in people's hands: "This is about getting the stimulus dollars out to projects quickly and providing jobs as soon as we can." In Section 9, I quantify the amount of public goods the government gave up as a result of this policy.

I am also interested in the effect of stimulus funded projects on the prices of projects funded from other sources. The linkage is again through the backlogs. Stimulus funded projects raise committed capacity, and thus we should expect higher prices for projects coming after them. However, these projects are funded by the states. There is then an externality that may have been overlooked by the states when receiving the stimulus funds. I quantify this effect in Section 9 as well. Table 1 shows the dollar amount of projects awarded by year, and for 2009 and 2010 it

Table 1: Projects by Year and Source

Year	Funds	Total Amount Awarded ($B)	Number of Projects	Avg Amount ($M)	Avg Length (days)
2006		2.666	598	4.457	167.7
2007		2.923	584	5.004	158.0
2008		3.551	656	5.413	142.8
2009	total	3.327	636	5.230	141.6
	other	2.336	560	4.172	131.0
	ARRA	.990	76	13.028	219.3
2010	total	3.141	648	4.846	155.5
	other	2.441	614	3.974	142.2
	ARRA	.700	34	20.589	395.2

includes a breakdown by source of funds. Stimulus funded projects accounted for 30% and 23% of the total dollar value in 2009 and 2010 respectively. Those projects were significantly larger on average (3 and 5 times as large, respectively) than projects from other sources. Also, the projects were longer on average. The effects on backlogs are then expected to be relatively larger.

4 Data and Procurement Process

In this section I describe the procurement process in California and the data I use. Additionally, I present some preliminary descriptive regressions to show that the data suggest the existence of both a backlog effect and unobserved heterogeneity. Although no causal effect can be claimed at this level, the regressions show that there is a positive correlation between a firm's backlog level and its bids, and a negative correlation with its individual participation decision. Also, high levels of backlogs are associated with a smaller number of bidders. These results suggest the presence of dynamic links between previously won projects and current outcomes.

I use highway procurement data from Caltrans from January 2000 through July 2011, from all 12 districts. For each project, I collect information from the publicly available bid summaries. The data correspond to road and highway maintenance projects' auctions, and include information on the bidding and award date, project characteristics (location of the job site, the estimated working days required for completion, and an engineer's estimate of the project cost in dollars), the number of planholders (potential bidders), the identities of the bidders, and their bids. Data from the same source, although for different time periods, have been used in previous studies by JP, Bajari, Houghton, and Tadelis (2006), and Krasnokutskaya and Seim (2011).

I complement Caltrans' data by constructing the backlog variable using previously won uncompleted projects. For each firm, at any point in time the backlog is defined as the amount of work in dollars that is left to do from previously won projects. Following Porter and Zona (1993) and JP, I assume a constant pace for the progression of work. Thus, for each firm and for every contract previously won, I compute the amount of work in dollars that is left to do by taking the initial size of the contract and multiplying by the fraction of time that is left until the project's completion date. To make this variable comparable across firms, I standardize it by subtracting the firm specific mean, and then dividing by the firm specific standard deviation.

I also collect data on firms' number of plants and their locations, and calculate the driving distance[13] to the job location. For projects with multiple locations I take the average across them, and for firms with multiple plans I consider the minimum distance, i.e., the distance to the closest plant. Finally, for each project I construct a measure of active bidders in the area by counting the number of distinct firms that bid in the same county over the prior year.

During the sample period I consider, Caltrans awarded 6686 projects with a total dollar value of $26.6 billion. In my analysis, I exclude projects where the main task is other than road and highway maintenance —such as landscaping, electrical work, or work on buildings— leaving 5753 auctions in my dataset.[14] Since I do not have auxiliary data to construct the backlogs prior to 2000, I drop the first year in my sample (589 auctions) in the estimation and use it to construct the backlogs at the beginning of 2001.

Caltrans uses a first-price sealed bid auction mechanism. The letting process involves the following steps. First, there is an initial announcement period of the project where few details are revealed (consisting of a short description of the project that includes the location, completion time, and a short list of the tasks involved). The advertising takes place three to ten weeks prior to the letting date, and depends on the size or complexity of the project. Second, interested firms may request a bid proposal document which includes the full description of the work to be done and

[13] I use Google Maps to get the distances. I also tried driving time instead of distance and results remained unchanged.

[14] Data for projects in the first semester of 2003 are missing. This accounts for 264 projects. Even though I cannot include these projects in my estimation, using auxiliary data I was able to reconstruct who the winner was, its bid and the number of working days. This is extremely important, because without these data I would have not been able to construct the backlog variable and would have had to discard data from 2000 to 2003 in the estimation.

project plans. Planholders may submit a sealed bid. Finally, on the letting day, bids are unsealed, ranked, and the project is awarded to the lowest bidder. The winning firm is awarded the contract within 30 days (for projects up to $200 million), or 60 days (for projects above $200 million).

While a firm cannot submit a bid without requesting the detailed specification, buying the project documents does not always result in a bid. On average, less than 45% of the planholders ends up submitting a bid. This may be rationalized from the fact that preparing the bid in this setting is costly. To bid on a project, a firm has to submit a bid bond (a predetermined percentage of the bid) and complete bid documents. The bid should include a detailed breakdown of costs by items (such as labor, mobilization, and materials). In most of the cases, firms have to subcontract parts of the work. This involves having to contact and negotiate with subcontractors. The process of reviewing the contract and submitting a bid is therefore costly, in terms of time and resources. Also, the bid preparation process I describe serves as a justification for an assumption I make later. It specifies that firms only learn their private project cost after they have decided to bid. Furthermore, I follow the standard convention in the literature and assume that firms know the identities of the other bidders. Krasnokutskaya and Seim (2011) justify this assumption in the highway procurement setting by the fact that it is common for bidders to share the same subcontractor in a given project, and thus it should be easy for them to learn who else is bidding.

In the data there are 1420 unique bidders. The vast majority only submits a bid once or just a few times in the 11 years of my sample period (52% of the firms submit at most 3 bids; 80% of the firms, at most 18 bids). On the other hand, there is a small group of firms that submit bids regularly throughout the period. Figure 1 shows a histogram for the number of bid submissions. Based on this fact, I consider two types of firms: regular and fringe. Regular firms are selected according to the following criteria: the top ten firms in terms of dollar value won who submit at least 300 bids in the sample period. At least one of the 10 regular firms participates in 77% of the auctions; regular firms won 23% of the projects which accounted for 27% of the total dollar value awarded.

Table 2 presents summary statistics for selected variables. On average, an auction attracts 14 planholders and close to 6 bidders. As was already mentioned, less than half of the planholders ends up submitting a bid. There is one regular bidder, on average, and the number ranges from 0 to 5. The average project has a duration close to 5 months. The variable $\frac{(\text{rank2}-\text{rank1})}{\text{rank1}}$ is also referred as "money left on the

Figure 1: Number of Bid Submissions

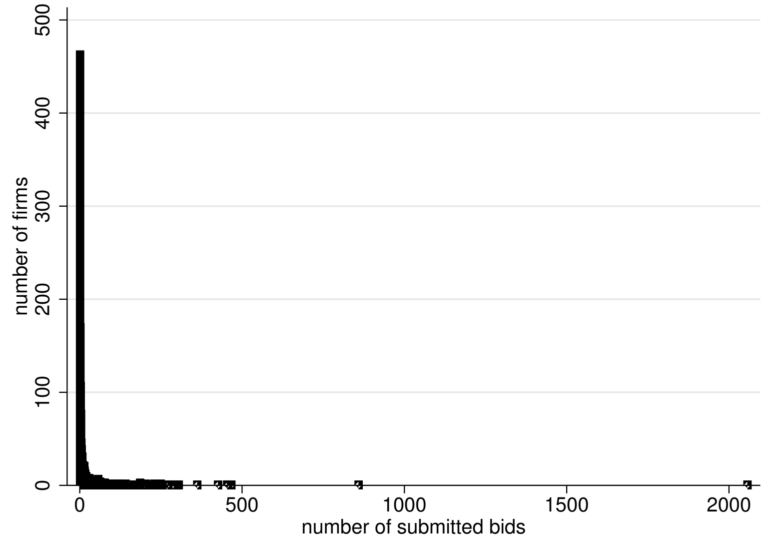


Table 2: Summary Statistics of Selected Variables

	Obs	Mean	Std. Dev.	Min	Max
Number of bidders	5753	5.67	3.11	1	23
Number of planholders	5753	14.36	8.49	2	50
bidders/planholders	5753	.44	.19	.02	1
Number of reg bidders	5753	.94	.86	0	5
log(estimate)	5753	13.90	1.41	11.17	19.79
Number of days	5753	127.67	178.19	7	1850
(rank2-rank1)/rank1	5625	.10	.12	0.00	1.49
(rank1-est)/est	5753	-.11	.25	-.83	2.30

table" in the auction literature. It measures the difference between the lowest and second lowest bid, as a fraction of the lowest bid, and gives an indication of the level of uncertainty or informational asymmetries in the market. On average, the second lowest bid is 10% higher than the winning bid. Table 3 provides a summary with a breakdown by number of bidders. Although "money left on the table" decreases with the number of bidders, it does not approach 0, suggesting that the magnitude of informational asymmetries may be large. This can also be seen from the relative difference of the winning bid to the engineer's estimate. It is on average -11%, and it is clearly decreasing in the number of bidders. It goes from being 24% above the estimate when there is only one bidder to 26% below it when there are 8 or more

Table 3: Summary Statistics of Selected Variables by Number of Bidders

# bidders		1	2	3	4	5	6	7	8+	All
# obs		128	556	822	899	791	713	535	1,309	5,753
log(estimate)	mean	13.47	13.91	13.82	13.92	13.96	13.99	14.08	14.02	13.90
	std dev	1.20	1.29	1.29	1.40	1.43	1.46	1.50	1.47	1.41
(rank2-rank1)/rank1	mean		0.19	0.12	0.10	0.08	0.08	0.07	0.08	0.10
	std dev		0.19	0.13	0.12	0.08	0.10	0.07	0.09	0.12
(rank1-est)/est	mean	0.24	0.05	-0.02	-0.08	-0.10	-0.15	-0.20	-0.26	-0.11
	std dev	0.41	0.26	0.24	0.23	0.21	0.19	0.19	0.18	0.25

bidders. Also note that larger projects, as measured by the engineers' estimate, seem to attract more bidders (without conditioning for any other variable).

I now turn to the descriptive regression results. Table 4 shows results for OLS regressions where the dependent variable is the log of bids. The first three columns include observations from all bidders, columns 4 and 5 include only bids from fringe bidders, and the last three columns include only regular bidders. The regressors include the (log) engineers' estimate, the (log) number of working days, the number of items involved in the project, the (log) distance between the firm and the project location, and the number of regular and fringe bidders in the auction. Some specifications include a dummy which equals one if the bidder is a regular firm, the firm's standardized backlog, the sum of standardized backlogs for regular firms participating in the auction, and the sum of standardized backlogs for regular firms not participating in the auction. All specifications also include time, district, and type of work fixed effects. All signs are as expected. Bigger projects (as proxied by the engineers' estimate) and greater distance to the job site are associated with larger bids, shorter projects, and a higher number of rivals are associated with lower bids.

As mentioned in the Introduction, backlogs may affect equilibrium bids via several channels. First, there is a direct effect: we expect the higher the firms' backlog, the higher the cost for completing a new project, and the higher the cost, the higher its bid. From Table 4 columns 6-8 we see that an increment of 1 standard deviation in a firm's backlog is associated with a 3% increase in its bid. The second channel is the strategic effect. That is, if rivals are constrained, even an unconstrained firm may bid less aggressively. We see that sum of (standardized) backlogs of rivals is positively correlated with the level of bids for both regular and fringe firms. For example, an increase in the backlogs of all 10 regular firms by 1 standard deviation is associated

Table 4: Bid Level Estimates (OLS)

	All bidders			Fringe bidders		Regular bidders		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
constant	.318***	.314***	.290***	.322***	.301***	.273***	.278***	.246***
	(.031)	(.030)	(.031)	(.034)	(.034)	(.065)	(.065)	(.066)
log(eng)	.973***	.973***	.973***	.973***	.973***	.973***	.973***	.974***
	(.005)	(.006)	(.005)	(.008)	(.006)	(.010)	(.010)	(.011)
log(days)	-.007**	-.007**	-.006*	-.004	-.004	-.017**	-.018**	-.018**
	(.003)	(.003)	(.003)	(.004)	(.003)	(.008)	(.008)	(.008)
items	.0005***	.0004***	.0004***	.0004***	.0004***	.0004***	.0004***	.0004***
	(.00006)	(.00006)	(.00006)	(.00008)	(.00007)	(.0001)	(.0001)	(.0001)
log(dist)	.022***	.023***	.023***	.022***	.022***	.025***	.025***	.025***
	(.002)	(.002)	(.002)	(.002)	(.002)	(.005)	(.004)	(.004)
# regular	-.023***	-.026***	-.025***	-.028***	-.027***	-.006	-.009	-.007
	(.002)	(.002)	(.002)	(.003)	(.003)	(.006)	(.005)	(.006)
# fringe	-.019***	-.018***	-.018***	-.018***	-.018***	-.021***	-.021***	-.020***
	(.0005)	(.0005)	(.0005)	(.0005)	(.0006)	(.001)	(.001)	(.001)
regular	-.006	-.006	-.006					
	(.005)	(.005)	(.005)					
std bl						.030***	.031***	.031***
						(.004)	(.004)	(.004)
sum std bl in auction		.023***	.024***	.023***	.024***		.014***	.015***
		(.001)	(.001)	(.002)	(.002)		(.004)	(.004)
sum std bl out auction			.004***		.004***			.006***
			(.0008)		(.0009)			(.001)
nobs	29606	29606	29606	24664	24664	4942	4942	4942
R^2	0.96	0.96	0.96	0.96	.96	.97	.97	.97

Dependent variable is log(bid). All regressions include time, district, and type of work dummies. Standard errors in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% level.

with a 6% higher bid for a fringe firm. For a given regular firm, conditional on its backlog, if other regular firms increase their backlog by 1 standard deviation, its bid is 6.3% higher.

Finally, the third channel is the competition effect. I show that higher backlogs are associated with a lower number of bidders (see Table 7), and we see in Table 4 that the lower the number of bidders (both regular and fringe) participating in an auction, the higher the equilibrium bids.

Table 5 shows the results for the OLS regression of (log) winning bid. Results are qualitatively similar to the previous table, except that some coefficients are now not statistically significant which may be due to the fewer number of observations.

Since I have 5164 auctions with a total of 29606 bids, and most of the auctions have more than one bid, I can exploit the panel structure of the data to control for unobserved (auction-level) heterogeneity observed by bidders when making their bidding decisions, but not observed by the econometrician. Table 6 shows the results for a random effects panel data model. Estimates of the coefficients for the observable variables do not vary significantly with respect to those from the OLS regressions, but the error variance from the unobserved heterogeneity accounts for 60 to 70% of the total error variance, supporting the existence of (auction-level) unobserved heterogeneity.

Now I show evidence that higher levels of backlogs are associated with less participation. In the first column of Table 7, I show the results from an OLS regression where the dependent variable is the number of bidders. The regressor potential firms is a proxy for the number of active firms in the area.[15] We observe that the sum of standardized backlogs is significant and enters with a negative sign, meaning that the level of backlogs is negatively correlated with the number of bidders. Since the dependent variable is a count variable, I also fit a Poisson regression model where the conditional mean is modeled as $E[N^B|x] = \exp(x\beta)$. Estimation results for the coefficients β are shown in the second column. Again, as the level of backlog increases, the equilibrium entry probability decreases. Since in the Poisson distribution the mean and variance are the same, I run a test for the goodness-of-fit. The large value for χ^2 is an indicator that the Poisson distribution is not a good choice. In the third column I show results for a negative binomial regression which is often more appropriate in cases of overdispersion, where α is the overdispersion parameter.[16] Although a likeli-

[15] It is constructed by counting the number of distinct firms that bid in the same county over the prior year.

[16] When the overdispersion parameter is zero, the negative binomial distribution is equivalent to a Poisson distribution.

Table 5: Bid Level Estimates: Winning Bid (OLS)

	All bidders			Fringe bidders		Regular bidders		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
constant	.110	.104	.097	.070	.068	.152	.156	.135
	(.073)	(.073)	(.074)	(.088)	(.089)	(.129)	(.129)	(.130)
log(eng)	.991***	.991***	.991***	.993***	.993***	.989***	.989***	.989***
	(.006)	(.006)	(.006)	(.007)	(.007)	(.010)	(.010)	(.010)
log(days)	-.015*	-.016*	-.016*	-.015	-.015	-.024	-.023	-.024
	(.008)	(.008)	(.008)	(.010)	(.010)	(.016)	(.016)	(.016)
items	.0006***	.0006***	.0006***	.0006***	.0006***	.0006**	.0006**	.0006**
	(.0001)	(.0001)	(.0001)	(.0002)	(.0002)	(.0003)	(.0003)	(.0003)
log(dist)	.011***	.012***	.012***	.012***	.012***	.012	.013	.013
	(.004)	(.004)	(.004)	(.005)	(.005)	(.010)	(.010)	(.010)
# regular	-.036***	-.040***	-.039***	-.041***	-.041***	-.017	-.021	-.019
	(.006)	(.006)	(.006)	(.007)	(.007)	(.013)	(.013)	(.014)
# fringe	-.038***	-.037***	-.038***	-.038***	-.038***	-.036***	-.036***	-.035***
	(.001)	(.001)	(.001)	(.002)	(.002)	(.003)	(.003)	(.003)
regular	.004	.003	-.003					
	(.011)	(.011)	(.012)					
std bl						.028***	.029***	.029***
						(.008)	(.008)	(.008)
sum std bl in auction		.022***	.022***	.021***	.021***		.011	.011
		(.004)	(.004)	(.006)	(.006)		(.010)	(.010)
sum std bl out auction			.001		.0004			.004
			(.002)		(.002)			(.004)
nobs	5164	5164	5164	4016	4016	1148	1148	1148
R^2	0.97	0.97	0.97	0.96	.96	.98	.98	.98

Dependent variable is log(bid). All regressions include time, district, and type of work dummies. Standard errors in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% level.

Table 6: Bid Level Estimates (Random Effects)

	All bidders			Fringe bidders		Regular bidders		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
constant	.391***	.386***	.365***	.370***	.352***	.408***	.402***	.372***
	(.064)	(.063)	(.064)	(.063)	(.064)	(.088)	(.088)	(.089)
log(eng)	.972***	.971***	.972***	.973***	.973***	.968***	.967***	.968***
	(.005)	(.005)	(.005)	(.005)	(.005)	(.007)	(.007)	(.007)
log(days)	-.003	-.003	-.003	-.002	-.002	-.022**	-.022**	-.022**
	(.008)	(.008)	(.008)	(.007)	(.008)	(.010)	(.010)	(.010)
items	.0005***	.0005***	.0005***	.0005***	.0004***	.0007***	.0006***	.0006***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
log(dist)	.016***	.016***	.016***	.016***	.016***	.013***	.013***	.013***
	(.001)	(.001)	(.001)	(.001)	(.002)	(.003)	(.003)	(.003)
# regular	-.030***	-.034***	-.032***	-.036***	-.035***	-.001	-.006	-.004
	(.005)	(.005)	(.005)	(.005)	(.005)	(.006)	(.008)	(.008)
# fringe	-.024***	-.023***	-.023***	-.023***	-.022***	-.023***	-.022***	-.021***
	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)	(.002)	(.002)
regular	-.015***	-.015***	-.015***					
	(.003)	(.003)	(.003)					
std bl						.013***	.031***	.031***
						(.002)	(.004)	(.004)
sum std bl in auction		.025***	.025***	.024***	.025***		.024***	.025***
		(.003)	(.003)	(.003)	(.003)		(.004)	(.004)
sum std bl out auction			.004***		.003*			.006***
			(.002)		(.002)			(.002)
nobs	29606	29606	29606	24664	24664	4942	4942	4942
$\sigma_u^2/(\sigma_u^2+\sigma_e^2)$	0.63	0.62	0.62	0.59	.59	.68	.69	.69

Dependent variable is log(bid). All regressions include time, district, and type of work dummies. Standard errors in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% level.

Table 7: Number of Bidders

	OLS	Poisson	Negative Binomial
constant	16.259***	3.404***	3.388***
	(.884)	(.112)	(.135)
log(eng)	-.742***	-.118***	-.119***
	(.067)	(.008)	(.010)
log(days)	.214**	.034***	.039**
	(.105)	(.013)	(.008)
items	.018***	.003***	.003***
	(.002)	(.0002)	(.0003)
potential firms	.0001	.00003	.00003
	(.0001)	(.0001)	(.0001)
sum std bl	-.244***	-.039***	-.039***
	(.022)	(.003)	(.003)
α			.070***
			(.005)
nobs	5164	5164	5164
R^2	0.15		
χ^2		4505	283

Dependent variable is the number of bidders. All regressions include time, district, and type of work dummies. Standard errors in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% level.

hood ratio test shows that the overdispersion parameter is significantly different from zero, the point estimates remain virtually unchanged.

As a last piece of evidence, Table 8 shows probit estimates for the participation decision of regular firms. The dependent variable is a binary variable indicating whether the firm participates in the auction. The firms' backlog coefficient is significant and enters with a negative sign. In fact, increasing a firm's backlog by 1 standard deviation makes the firm 15% less likely to participate, and a firm with a standardized backlog equal to 2 is 55% less likely to participate than a firm with a standardized backlog equal to -2.

5 The Model

This section describes the model I take to the data. The model I develop here is based on JP's repeated first-price bidding game in which firms' costs are intertemporally linked through their backlogs. I depart from it in two critical ways. I allow for auction level unobserved heterogeneity and endogenous participation.

Table 8: Regular Bidders Participation Decision

	(1)	(2)
constant	-.784***	-.729***
	(.168)	(.169)
log(eng)	.168***	.167***
	(.012)	(.012)
log(days)	-.041**	-.041**
	(.020)	(.020)
items	-.0007*	-.0007*
	(.0003)	(.0003)
log(dist)	-.602***	.603***
	(.010)	(.010)
potential bidders	-.001***	-.001***
	(.0002)	(.0002)
std bl	-.069***	-.073***
	(.010)	(.010)
sum std bl rivals		-.015***
		(.004)
nobs	51640	51640

Probit estimates for the decision to participate (regular firms only). All regressions include time, district, and type of work dummies. Standard errors in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% level.

Time is discrete with an infinite horizon, $t = 1, 2, \ldots$. I consider two types of risk-neutral firms. My main focus is on regular firms which are long-lived and stay in the game forever. The number of regular firms is fixed and known, and I denote it by N^{reg}. Fringe firms are short-lived: they participate in an auction and then die. This modeling choice reflects my finding that a small group of firms submits bids regularly in the sample period and a large group of firms submits bids just a few times (see Section 4). For the latter group, I do not have enough observations to describe their dynamic behavior.

Each project is associated with characteristics (Y, W) which are drawn independently and identically from the distribution function $F_{Y,W}(\cdot)$ with finite support $S_{Y,W}$.[17] These characteristics include all the information about the project such as

[17]Throughout the paper, random variables are in uppercase and their realizations in lowercase; vectors are in boldface. I denote the cumulative distribution function of a latent random variable X by $F_X(\cdot)$, and that of an observable random variable X by $G_X(\cdot)$. I use lowercase for their associated densities or probability mass if the random variable is discrete. Later in the paper, I use lowercase boldface to denote some matrices.

the size and duration of the project, its location, the quality of the existing road, etc. The distinction between Y and W is that the latter represents auction specific unobservable factors, i.e., factors observed by the firms participating in the auction but not by the econometrician. I assume that the unobserved factors can be summarized as a scalar index. I assume that future project characteristics are not known to the firms at time t, but the distribution function $F_{Y,W}(\cdot)$ is common knowledge.

The existence of unobserved auction heterogeneity is an important feature of my model. It is also a critical difference from JP. As has been documented by Krasnokutskaya (2011) in the highway procurement setting, costs can be substantially affected by local conditions, such as elevation, curvature, traffic, age, or quality of the existing road. These conditions are included in the project plans and documentation and hence observed by the firms, but typically are not observed by the econometrician.[18]

I assume an independent private values (IPV) setting (symmetric conditional on project and firms' characteristics). This assumption is justified by the fact that in road/highway maintenance projects, the project itself is precisely specified, hence bidders can accurately predict their own costs (as opposed to a common-value setting). The source of variation in private costs comes from firms having different opportunity costs for their own resources or differences in the input prices they face.

I consider a general non-separable structure for the private costs. For regular bidder i in auction t, the cost of completing the project is given by

$$C_{it} = c(\tilde{C}_{it}, Y_t, W_t, s_{it}) \tag{1}$$

where $\tilde{C}_{it}$ is the private type, and s_{it} is a vector of state variables for bidder i.[19] Bidder i's state vector, s_i, includes a list of the sizes of all uncompleted projects won by i in the past and the time to complete each of them. I assume that the vector of all regular bidders' state variables, denoted by $\boldsymbol{s}$, is observed by all bidders and the econometrician. Note that bidder i observes all components of (1) separately and not just the realization of C_{it}. Krasnokutskaya (2011) considers a similar cost structure, although she imposes a linear functional form on (1), an assumption that is crucial for identification in her setting. Li, Perrigne, and Vuong (2000) also consider a similar structure where bidder's costs are composed of common and individual factors. However, bidders do not observe the realization of the common factor separately from

[18] Additionally, it is common for firms to send engineers to the location to assess these factors before submitting a bid.

[19] Implicitly I am assuming that all regular firms have the same cost function $c(\cdot)$. This is not necessary and it is straightforward to allow for a different cost function for each firm.

the entire realization of their costs. Also, critical for identification in their setting is the linear functional form for (1).

Similarly, for fringe bidder j at auction t we have

$$(2) \quad C_{jt} = c^f(\tilde{C}_{jt}, Y_t, W_t).$$

From now on I drop the subscript t for simplicity. The cost of a regular bidder i is drawn from the continuous conditional distribution $F_{C|YWs}(\cdot|\cdot)$ with support $[\underline{C}(Y,W,s), \bar{C}(Y,W,s)]$, and the cost of a fringe bidder is drawn from the continuous distribution function $F^f_{C|YW}(\cdot|\cdot)$, with support $[\underline{C}^f(Y,W), \bar{C}^f(Y,W)]$.

The only restriction I require on (1) (and an analogous one for fringe firms) is the following

Assumption 5.1 *$E[C_i|y, W = l, s] < E[C_i|y, W = m, s]$ for $l < m$, all i, y, s.*

The previous assumption is a stochastic monotonicity condition on the unobservable. This is a reasonable assumption if we think of the unobserved factor as some measure of the project's quality. Sufficient conditions (omitting dependence on observed variables for simplicity) are, for example, that the conditional distributions of the private costs satisfy $F_{C_i|W}(x|W = l) > F_{C_i|W}(x|W = m)$ for $l < m$, all x; or the assumptions in Krasnokutskaya (2011), namely, $\tilde{C} \perp W$ and $C = c(\tilde{C}, W)$ strictly increasing in both arguments. These assumptions are stronger than 5.1 and not necessary in my setting.

5.1 The stage game

The stage game is based on the actual procurement process in California. At every period t the buyer offers a single contract for sale. The timing of the stage game is as follows:

1. Advertising period: firms observe (Y^A, W^A) where Y^A is a subset of Y and W^A is a noisy signal of W.

2. Based on (Y^A, W^A) and backlogs, firms simultaneously decide to request the documents containing the full details of the project. I assume that the number of planholders and their identities is not public information.

3. If the firm has requested the documents, it learns (Y, W) and gets a draw of the bid preparation cost, κ_i^B, which I assume is iid across firms and time. The bid preparation cost is private information.

4. Based on (Y, W), κ_i^B, and backlogs, the firm decides whether to prepare a bid (before learning its private project cost).

5. If the firm decides to prepare a bid, it learns its private project cost, c_i. I assume that while learning the private cost, the firm also learns the identity of its rivals.[20] The firm then submits a bid (no reserve price).

6. The buyer awards the contract to the low bid firm at a price equal to its bid.

Consistent with the data, in the advertising period firms do not observe the complete set of characteristics but just the few characteristics that are advertised and a noisy signal about the unobserved project level factor.[21] In this setting, κ_i^B is a signal acquisition cost as in Levin and Smith (1994) (which has been used in empirical applications in Bajari and Hortacsu (2003), Athey, Levin, and Seira (2011), Athey, Coey, and Levin (2011), Krasnokutskaya and Seim (2011), and Groeger (2010)) and *not* a bid preparation cost as in Samuelson (1985). In the latter case, firms first observe their project cost and then decide whether to incur the cost to prepare their bid.[22]

5.2 Participation Decisions

In my model, participation is endogenous in the sense that I allow project characteristics (observed and unobserved) to affect the number of bidders.[23] That is, firms may decide to participate in an auction after learning the project characteristics but prior to learning their private cost. Although taking to the data a full structural model that includes individual participation decisions (such as the one described in Section 5.1) jointly with the bidding decision would improve the efficiency of my estimates, it is beyond the scope of this paper. Instead, I follow HHS and use reduced form

[20] The same assumption has been used in Krasnokutskaya (2011), Krasnokutskaya and Seim (2011), Athey, Levin, and Seira (2011), and Athey, Coey, and Levin (2011).

[21] For example, if the unobserved factor is the quality of the existing road, firms may know that in general the quality of the road to be repaired is "bad," but they do not know the exact quality level for the specific section of the road involved in the project.

[22] Li and Zheng (2009) jointly model the entry and bidding decisions under both Levin and Smith- and Samuelson-type of models. Using highway procurement data, they find that the Levin and Smith model fits the data better. Roberts and Sweeting (2010) and Marmer, Shneyerov, and Xu (2011) take an alternative route, and instead of considering the two polar cases they allow bidders to have an imperfectly informative signal about their value prior to deciding whether to pay the entry cost. As signals become more (less) informative, the model approaches the assumption in Samuelson (Levin and Smith).

[23] Although I will use the term participation equations for both the decision to become a planholder and the decision to submit a bid, hereafter when I say that a firm participates in an auction I am referring to a firm that has payed the signal acquisition cost and becomes an active bidder.

equations which are based on the structure of the model described in the previous section. This is enough to allow me achieve identification of the objects of interest.

As mentioned earlier, endogenous participation is one chief deviation that I make from JP. In their model, the number of bidders is exogenously determined: regular firms always bid, and fringe firms' participation is taken as exogenous. They also make the assumption that the researcher may not observe all bids from regular bidders due to a reserve price. That is, bids that are above the reserve price are rejected and not recorded.[24] In extensive conversations with engineers at Caltrans, they confirmed that there is no reserve price (publicly announced or not), and that all bids received are recorded.[25] This is consistent with the no reserve price assumption made by other studies that use Caltrans data (see Section 4 for references).

The setup and assumptions in this section closely follow HHS with two differences. I consider two participation equations, one for the number of planholders and the other for the number of bidders. Additionally, I relax the structural assumption that requires the observable variables in each of the equations to be a "sufficient" statistic for the unobservable, W. In the auction literature, Campo, Perrigne, and Vuong (2003) and Guerre, Perrigne, and Vuong (2009) also use the latter assumption to solve the unobserved heterogeneity problem. This may be a strong assumption on the structure of the model, and there are at least two reasons why it might not hold in many applications. First, other unobservable factors may enter into the endogenous outcome determination. Second, if the unobservable factor is realized after the endogenous outcome decision that the researcher is exploiting, the endogenous outcome may rely only on a noisy signal of the unobserved heterogeneity.

Let N^A denote the number of planholders, i.e., the number of firms that show interest in the advertising period and request the documentation, and let N^B denote the number of bidders. According to the timing of the stage game, I write

$$N^A = \phi^A(Y^A, W^A, \boldsymbol{s}) \tag{3}$$

$$N^B = \phi^B(Y, W, \boldsymbol{\kappa^B}, \boldsymbol{s}) \tag{4}$$

where $W^A = \omega^A(W, u)$ is a *noisy* signal of W, u is an unobserved (both to the firms and the econometrician) error,[26] $\boldsymbol{\kappa^B}$ is the vector of signal acquisition costs, and

[24] JP claim that they observe that 12% of all bids are above the reserve price. They consider that as erroneous and exclude them from their analysis.

[25] Caltrans may reject a bid if it considers that the firm is not responsive or that it has not met the qualification requirements. Even in this case, the bid is recorded and marked as rejected.

[26] Since W is a discrete random variable, u is in fact a misclassification error, and W^A is said to be a misclassified measurement of W.

both $\phi^A(\cdot)$ and $\phi^B(\cdot)$ are unknown functions. Note that I use the term *reduced form* in a very precise way. Equations (3) and (4) represent the relationships between the respective participation outcomes and auction characteristics and firms' types implied by the underlying structural model.

I further assume that the vector of signal acquisition costs, $\boldsymbol{\kappa}^{\boldsymbol{B}}$ and W enter (4) as a scalar index $W^B = \omega^B(W, \boldsymbol{\kappa}^{\boldsymbol{B}})$. This may be a restrictive assumption but it is typical in the literature to reduce the dimensionality of the unobservables to a scalar. Nevertheless, there is a sense in which the scalar assumption here is weaker than in the traditional control function approaches used in auctions. This stems from the fact that I am allowing the unobserved factor W^B to differ from the unobserved factor entering the bidding decision (i.e., the *outcome* equation).

Then I can rewrite the equation for the number of bidders as

$$N^B = \phi^B(Y, W^B, \boldsymbol{s}). \tag{5}$$

I make the following assumptions:

Assumption 5.2 *W^A is independent of $(Y^A, \boldsymbol{s})$; and W^B is independent of $(Y, \boldsymbol{s})$.*

Assumption 5.3 *For all $(y^A, \boldsymbol{s})$, the support of $N^A|(y^A, \boldsymbol{s})$ is a finite convex subset of $\mathbb{Z}_+$; for all $(y, \boldsymbol{s})$, the support of $N^B|y, \boldsymbol{s}$ is a finite convex subset of $\mathbb{Z}_+$*

Assumption 5.4 *ϕ^A is strictly increasing in W^A; ϕ^B is strictly increasing in W^B.*

Assumption 5.2 is not necessary but greatly simplifies the proofs that follow.[27] Assumption 5.3 seems harmless since the number of firms in the market is bounded from below by zero and from above by the total number of firms, for example, in the country. On the other hand, the strict monotonicity of the functions $\phi^A(\cdot)$ and $\phi^B(\cdot)$ on W^A and W^B, respectively, is a strong restriction since it requires that both W^A and W^B be discrete.[28] This assumption is key in my procedure since I rely on it to invert the relationships and recover the unobserved W^A and W^B, and is a typical assumption in other nonparametric control function strategies (see Chesher (2003), Imbens and Newey (2009)). Although the discreteness of W^A and W^B is a strong

[27] The proofs still hold conditional on $(Y^A, \boldsymbol{s})$ and $(Y, \boldsymbol{s})$

[28] The same kind of discreteness of the unobservable arises in Campo, Perrigne, and Vuong (2003) for the same reasons I articulate.

assumption, it is common in practice to assume a discrete support for unobserved factors to nonparametrically approximate its distribution.[29][30]

Since W^A is a misclassified measurement of W, both variables share the same support, which will be denoted by $\{w_k\}_{k=1}^{K}$ with K unknown but finite from Assumption 5.3. The support of W^B is denoted by $\{w_k^B\}_{k=1}^{K^B}$ where, again, K^B is unknown but finite. Let the associated probabilities be $\{p_k^A\}_{k=1}^{K}$ and $\{p_k^B\}_{k=1}^{K^B}$, respectively. The identification result in Section 6.2 requires that $K \leq K^B$. In what follows, for ease of exposition and without loss of generality, I assume $K^B = K$.[31]

Discussion. Although I have started from a structural model of participation decisions (Section 5.1) and based the reduced form equations on it, the procedure developed in this paper does not require it. In fact, one can replace equations (3) and (4) by any other two endogenous outcome equations, for example, the number of regular bidders and the number of fringe bidders. The identification proof does not rely on the structural participation model. It just requires that the researcher be able to obtain two potentially noisy measurements of W. This is an advantage of the procedure, but at the same time there is a limitation. By relying on reduced form equations, I cannot analyze some kinds of counterfactuals like changes in the auction rules not captured by changes in the variables entering equations (3) and (5). This criticism is also shared by other control function approaches that rely on a reduced form equation (e.g., HHS and Roberts (2011)).[32] Nevertheless, there are interesting and important questions that can still be answered such as the ones raised in this paper.

5.3 Bidding Decision

Conditional on participation, bidders have to optimally choose their bids. Let $\mathcal{N}$ denote the set of bidders at the auction. It is enough that $\mathcal{N}$ includes the identities of the regular bidders and just the number of fringe bidders (not their identities).

[29] In fact, it is not obvious that specifying a know parametric distribution for a continuous unobserved factor, is better than assuming a discrete support, since economic theory offers little guidance on the actual functional form of the distribution. In fact, the choice of a particular distribution of unobservables is usually justified on the grounds of familiarity, ease of manipulation, and considerations of computational cost. See Heckman and Singer (1984) for a discussion in the context of duration models.

[30] In the application, I consider 20 points in the supports of W^A and W^B.

[31] The identification proof still holds, but I need to replace the matrix inverse operator with generalized inverses for non-square matrices.

[32] Although Roberts does not use a participation reduced-form equation, he uses a reserve price reduced form equation subject to the same criticism. Changes affecting the way the reserve price is set cannot be considered in counterfactuals.

Consistent with the timing of the stage game and equation (5), the set of bidders $\mathcal{N}$ is a function of $(Y, W, \boldsymbol{s}, \boldsymbol{\kappa^B})$, and $|\mathcal{N}| = N^B$. At the bidding stage $(y, w, \mathcal{N}, \boldsymbol{s})$ is observable to all bidders. Let b_i denote bidder i's bid.

Let $x = (z, \tau)$ be the size and duration of the project (which is in fact a subset of the vector y) with finite support S_X. Let S denote the support of $\boldsymbol{s}$. Then, the transition function of the regular firms' state variable $\omega : S_X \times S \times \{1, \ldots, n\} \to S$ is a deterministic function of the contract size, length, the state variables and the identity of the winner. This function updates the backlogs of each of the regular firms as follows. Since I do not observe the pace at which projects are completed, I assume a constant progression over time.[33] That is, at every point in time, an equal share of the project is completed. As time advances one period, for each firm all previously won and uncompleted projects decrease their size proportionally, and the length is reduced by one unit. If firm i is the current winner, then the project size and length is added as the first element of s_i, otherwise $(0,0)$ is added. Mathematically, I can write the i-th component of the transition function as

$$\omega_i(x, \boldsymbol{s}, j) = \begin{cases} \left((z, \tau), \left(\frac{\max(\tau_i^l - 1, 0)}{\tau_i^l} z_i^l, \max(\tau_i^l - 1, 0)\right)_{l=1}^{\bar{\tau}-1}\right) & \text{if } j = i \\ \left((0, 0), \left(\frac{\max(\tau_i^l - 1, 0)}{\tau_i^l} z_i^l, \max(\tau_i^l - 1, 0)\right)_{l=1}^{\bar{\tau}-1}\right) & \text{if } j \neq i \end{cases}$$

where $\bar{\tau}$ is the maximum length of a project.

Bidders discount the future with a common discount factor $\beta \in (0, 1)$. The discount factor is known to the econometrician and bidders and constant over time.

Conditional independence of contract characteristics and cost realizations is a crucial assumption that allows me to adopt a Markov dynamic decision process. I consider a Markov-perfect equilibria concept (and restrict to symmetric strategies). This means that the equilibrium strategies do not depend on time. Let $b(c_i, y, w, \mathcal{N}, s_i, s_{-i})$ be i's strategy, and let b_{-i} denote the strategy profile of rivals.

Since the outcome of the auction affects not only current profits but also the firm's backlog, firms choose their bids so as to maximize the expected discounted value of future profits. The discounted sum of future expected payoffs for regular bidder i can

[33] The same assumption is used in JP, Porter and Zona (1993), Li and Zheng (2009), and Bajari, Houghton, and Tadelis (2006).

be written in value function form as

$$
\begin{aligned}
(6) \quad M_i(y,w,\mathcal{N},\boldsymbol{s},c,b_{-i}) \quad = \quad & 1\{i\in\mathcal{N}\}\max_b\Big\{(b-c)\Pr(i \text{ wins}|b,y,w,\mathcal{N},s_i,s_{-i}) \\
& +\beta\sum_{j\in\mathcal{N}}\Pr(j \text{ wins}|b,y,w,\mathcal{N},s_i,s_{-i}) \\
& \times \mathrm{E}_{YW\mathcal{N}}\Big[\int M_i(y',w',\mathcal{N}',\omega(x,\boldsymbol{s},j),c',b_{-i}) \\
& \times f(c'|y',w',\mathcal{N}',\omega_i(x,s,j))dc'\Big]\Big\} \\
& +1\{i\notin\mathcal{N}\}\beta\sum_{j\in\mathcal{N}}\Pr(j \text{ wins}|y,w,\mathcal{N},\boldsymbol{s}) \\
& \times \mathrm{E}_{YW\mathcal{N}}\Big[\int M_i(y',w',\mathcal{N}',\omega(x,\boldsymbol{s},j),c',b_{-i}) \\
& \times f(c'|y',w',\mathcal{N}',\omega_i(x,\boldsymbol{s},j))dc'\Big]
\end{aligned}
$$

Let $G(\cdot|y,w,\mathcal{N},s_i,s_{-i})$ denote the distribution function of equilibrium bids of bidder i with state $(y,w,\mathcal{N},s_i,s_{-i})$ with associated density $g(\cdot|y,w,\mathcal{N},s_i,s_{-i})$. In the same way, I write the distribution function of equilibrium bids of a fringe bidder with state $(y,w,\mathcal{N},\boldsymbol{s})$ by $G_f(\cdot|y,w,\mathcal{N},\boldsymbol{s})$ (with associated density $g_f(\cdot|y,w,\mathcal{N},\boldsymbol{s})$). In an abuse of notation, I will write $G(\cdot|y,w,\mathcal{N},s_i,s_{-i})$ where i can take value f (and it is understood that we refer to $G_f(\cdot|y,w,\mathcal{N},\boldsymbol{s})$). Doing so, allows me to write the probability that bidder i wins when she submits a bid equal to b as

$$
\Pr(i \text{ wins}|b,y,w,\mathcal{N},s_i,s_{-i}) = \prod_{j\in\mathcal{N},j\neq i}[1-G(b|y,w,\mathcal{N},s_j,s_{-j})].
$$

The probability that bidder i assigns to the event that bidder j wins when i bids b, is given by

$$
\Pr(j \text{ wins}|b,y,w,\mathcal{N},s_i,s_{-i}) = \int_{\underline{b}}^{b} g(x|y,w,\mathcal{N},s_j,s_{-j})\prod_{l\in\mathcal{N},l\neq i,j}[1-G(x|y,w,\mathcal{N},s_l,s_{-l})]dx.
$$

Finally,

$$
\Pr(j \text{ wins}|y,w,\mathcal{N},\boldsymbol{s}) = \int_{\underline{b}}^{\bar{b}} g(x|y,w,\mathcal{N},s_j,s_{-j})\prod_{l\in\mathcal{N},l\neq j}[1-G(x|y,w,\mathcal{N},s_l,s_{-l})]dx
$$

is the probability that bidder i assigns to the event that bidder j wins when i does not participate in the auction.

It is convenient to write the maximization problem at the beginning of a period, prior to the realization of the private cost and prior to the realization of the contract characteristics. Therefore, I define the ex ante value function as the value function evaluated before Y, W, $\mathcal{N}$, and private costs are realized. I write it as

$$V_i(\boldsymbol{s}, b_{-i}) = \mathrm{E}_{YW\mathcal{N}}\Big[\int M_i(y, w, \mathcal{N}, \boldsymbol{s}, c, \beta_{-i}) f(c|y, w, \mathcal{N}, s_i) dc\Big]. \tag{7}$$

Due to the Markov structure of the problem, equation (7) can be written recursively as follows, dropping the dependence on rivals' bidding strategies for notational simplicity,

$$\begin{aligned} V_i(\boldsymbol{s}) \;\; = \;\; & E_{YW\mathcal{N}}\Big[1\{i \in \mathcal{N}\} \int \max_b \Big\{(b - c)\Pr(i \text{ wins}|b, y, w, \mathcal{N}, s_i, s_{-i}) \\ & +\beta \sum_{j \in \mathcal{N}} \Pr(j \text{ wins}|b, y, w, \mathcal{N}, s_i, s_{-i}) V_i(\omega(x, \boldsymbol{s}, j))\Big\} f(c|y, w, \mathcal{N}, s_i) dc \\ & +1\{i \notin \mathcal{N}\}\beta \sum_{j \in \mathcal{N}} \Pr(j \text{ wins}|y, w, \mathcal{N}, s) V_i(\omega(x, \boldsymbol{s}, j))\Big] \end{aligned} \tag{8}$$

Remember that fringe bidders enter in an auction and then die, therefore for fringe bidder i the ex ante payoff is equal to the ex ante expected period payoff,

$$E_{YW\mathcal{N}}\Big[\int \max_b \Big\{(b - c)\Pr(i \text{ wins}|b, y, w, \mathcal{N}, \boldsymbol{s}) f^f(c|y, w, \mathcal{N}) dc\Big]. \tag{9}$$

It should be clear from the previous equation that fringe bidders assign no value to the future.

Equilibrium existence. The equilibrium existence discussion in JP still applies to the present model. Note that the realizations of both W and $\mathcal{N}$ are observed by the firms at the time of bidding.

6 Nonparametric Identification

In this section I show the nonparametric identification of the structural elements of the model. My underlying identification strategy is similar in spirit to the control function approach (see Chesher (2003) and Imbens and Newey (2009), and HHS or Roberts (2011) for applications to auctions). The control function approach relies on an equation that relates an endogenous observed outcome to the unobserved factor. With a strict monotonicity assumption, one can invert the relationship and use it to control for the unobserved factor directly. I depart from this method by allowing for an "imperfect" control function. That is, I do not require the observables in the relationship to be a sufficient statistic for the unobserved heterogeneity. Rather, I exploit features of the procurement setting that provide a second "imperfect" control function. The information obtained from these two noisy controls then resembles a measurement error problem where we have access to multiple measurements. I attain identification of the distribution functions of equilibrium bids using the result in Hu (2008) for nonlinear models with misclassification error. The control functions I use are derived from the reduced-form participation equations (3) and (5) in Section 5.2.

The main result states that $F_{C|YWs}(\cdot|\cdot)$, $F^f_{C|YW}(\cdot|\cdot)$, and $F_W(\cdot)$ are nonparametrically identified. The proof proceeds in three parts. First, I show that the reduced form participation equations are identified. Second, I show that the conditional distribution functions of equilibrium bids are identified. Finally, I show that the value function is identified and, using the first order condition at the bidding stage, I show that the conditional distribution functions of private costs are identified.

6.1 Participation Decisions

Here I show that the participation equations are nonparametrically identified. The results in this section follow from HHS.

First, it is useful to note that under assumptions 5.2-5.4 I can rewrite the *reduced form* equations without loss as

$$N^A = \zeta^A(Y^A, \boldsymbol{s}) + W^A$$

$$N^B = \zeta^B(Y, \boldsymbol{s}) + W^B$$

where $\zeta^A(\cdot)$ and $\zeta^B(\cdot)$ are unknown integer valued functions (see HHS for a proof).

In the data there are a few auctions that attracted zero bidders. For those auctions, I do not observe the characteristics Y, and hence discard them from the estimation. This results in sample selection, in particular, truncation from below at 1 in the reduced form equation for the number of bidders. HHS provide conditions for identification when the sample is truncated both from below and from above. The conditions I need for one-sided truncation are weaker. In particular, I require that there is enough variation in the observables so that sometimes N^B is not truncated:

Assumption 6.1 $\exists (y, \boldsymbol{s})'$ *such that min supp*$\{N_B|(y, \boldsymbol{s})'\} = 2$

Theorem 6.1 *Under assumptions 5.2-5.4 and 6.1, from the joint distribution of the selected sample of observables the following objects are identified* *(i)* $\zeta_A(\cdot)$ *and* $\zeta_B(\cdot)$ *up to a location normalization,*

(ii) the number of points in the support of W^A *and* W^B*, i.e.,* K *and* K^B*,*

(iii) the joint distribution of W^A *and* W^B*.*

Proof Let $(\tilde{N}^A, \tilde{N}^B, \tilde{Y})$ denote the selected sample. That is, $(\tilde{N}^A, \tilde{N}^B, \tilde{Y}) = (N^A, N^B, Y)$ is observed only if $N^B \geq 1$. Remember that the support of W^A is given by $\{w_k\}_{k=1}^{K}$ and the support of W^B is given by $\{w_k^B\}_{k=1}^{K^B}$.

Without loss, I impose the following normalization: $w_k = k$ for $k = 1, \ldots, K$ and, similarly, $w_k^B = k$ for $k = 1, \ldots, K^B$.

From the joint distribution of $(\tilde{N}^A, \tilde{N}^B)$ conditional on $(y, \boldsymbol{s})'$, K and K^B are identified and are equal to the number of points in the support of $\tilde{N}^A|(y^A, \boldsymbol{s})'$ and of $\tilde{N}^B|(y, \boldsymbol{s})'$, respectively. Also, the joint distribution of W^A and W^B, denoted by $p_{ij}^{AB} = \Pr(W^A = i, W^B = j)$, is identified.

Now, the distribution of $\tilde{N}^A|(y^A, \boldsymbol{s})$ reveals $\zeta_A(\cdot)$ for all $(y^A, \boldsymbol{s})$ in the support $\tilde{Y}^A \times S$, just set $\zeta_A(y^A, \boldsymbol{s}) = \max \operatorname{supp}\{\tilde{N}_A|(y^A, \boldsymbol{s})\} - K$. By the same argument, the distribution of $\tilde{N}^B|(y, \boldsymbol{s})$ reveals $\zeta_B(\cdot)$ for all $(y, \boldsymbol{s})$ in the support $\tilde{Y} \times S$. $\square$

6.2 Distributions of Equilibrium Bids

In this section I show that, despite the fact that the econometrician does not observe W, the conditional distribution of equilibrium bids, $G(\cdot|y, w, \mathcal{N}, s_i, s_{-i})$, and the distribution of unobserved heterogeneity, $F_W(\cdot)$, are nonparametrically identified. In my setting, the underlying problem is a finite mixture problem.[34] This section relies on recent results for nonlinear models with misclassification error by Hu (2008). All the results in this section hold conditional on the project's observable characteristics and state variables. Thus, for notational simplicity I omit conditioning on them.

Contrary to previous studies, the obstacle to identifying the distribution of equilibrium bids in my model stems from fact that neither control function (3) or (5) allows me to recover the latent unobserved factor W. If one were able to do so, one can just directly control for W in the bid distribution functions after inverting the control function. This is the strategy followed, for example, by HHS, Campo, Perrigne, and Vuong (2003), and Guerre, Perrigne, and Vuong (2009). The reason I

[34] Some recent work on nonparametric identification of finite mixture models includes Mahajan (2006), Lewbel (2007), Chen, Hu, and Lewbel (2008a), Chen, Hu, and Lewbel (2008b), Chen, Hu, and Lewbel (2009), Hu (2008), Chen, Hong, and Tamer (2005), Hall and Zhou (2003), Kitamura (2003), and Henry, Kitamura, and Salanie (2011).

cannot recover the latent unobserved factor W is the existence of other unobserved factors, namely u and the vector $\boldsymbol{\kappa^B}$, that *contaminate* the control functions.

In the production function estimation context,[35] Ackerberg, Benkard, Berry, and Pakes (2007) relax the scalar unobservable assumption. For expositional simplicity, they proceed by assuming that a 2-dimensional unobservable enters in the endogenous outcome equation (the investment equation in their case). One of the unobserved factors is the one that needs to be controlled for in the outcome equation (the production function) and the other is treated as a nuisance parameter. More structure is required, so they add a second observed endogenous outcome equation, which also depends on the two unobserved factors. Assuming that the system of endogenous outcome equations is a bijection in the unobserved factors, they can invert the system and recover the unobserved factor of interest. In that case, the traditional control function approach still applies. My approach requires a second equation too, but I cannot assume that the latent unobserved factor can be recovered by inverting the system of equations. Instead, I just assume that I can recover two noisy measures of it (where the measurement errors have to satisfy certain assumptions) and apply the results from Hu (2008) to identify the distributions of equilibrium bids conditional on the unobserved heterogeneity. In particular, I make the following assumptions regarding W^A and W^B:

Assumption 6.2 *Conditional on* $(Y, W, \boldsymbol{s})$, $(C, U, \boldsymbol{\kappa^B})$ *are jointly independent.*

Define the K-dimensional square matrix[36] $\mathbf{g_{W^B,W^A}} \equiv [\Pr(W^B = i, W^A = j)]_{i,j}$ $i, j = 1, \ldots, K$.

Assumption 6.3 *Rank*$(\mathbf{g_{W^B,W^A}}) = K$

A first implication of assumption 6.2 is that W^A and W^B do not contain any useful information on the private values C beyond that of the true value of W. In particular, the signal acquisition costs are assumed to be independent of the project completion cost. A second implication is that I require the misclassification error in W^A, U, to be independent of the bid preparation costs once I condition on W and the state variables. Assumption 6.3 requires some statistical dependence between W^A and W^B although it does not necessarily require correlation. This is natural since both are meant to be informative signals of W. An advantage of the latter assumption is that it can be tested from the data.

[35]See Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Fraser (2004)

[36]The identification result requires $K \leq K^B$. Remember that without loss of generality I assumed $K = K^B$.

Theorem 6.2 *Under assumptions 5.1-5.4 and 6.1-6.3, $G(\cdot|y, w, \mathcal{N}, s_i, s_{-i})$ and $F_W(\cdot)$ are nonparametrically identified*

Proof See the Appendix. $\square$

6.3 Distributions of Private Values

I start by deriving the first-order condition (FOC) at the bidding stage. For a regular bidder i with private project completion cost c_i, the FOC corresponding to the optimal bid is given by

$$
\begin{aligned}
(10) \quad c_i &= b - \frac{1}{\sum_{j \neq i} h(b|y, w, \mathcal{N}, s_j, s_{-j})} \\
&\quad + \beta \sum_{j \neq i} \frac{h(b|y, w, \mathcal{N}, s_j, s_{-j})}{\sum_{l \neq i} h(b|y, w, \mathcal{N}, s_l, s_{-l})} [V_i(\omega(x, \boldsymbol{s}, i)) - V_i(\omega(x, \boldsymbol{s}, j))] \\
&\equiv \xi(b|y, w, \mathcal{N}, s_i, s_{-i}, h, V_i)
\end{aligned}
$$

where $h(\cdot|y, w, \mathcal{N}, s_i, s_{-i}) = \frac{g(\cdot|y,w,\mathcal{N},s_i,s_{-i})}{1-G(\cdot|y,w,\mathcal{N},s_i,s_{-i})}$ is the hazard function of bids. The FOC relates firm i's private information[37] c_i to its bid, the distribution and density functions of equilibrium bids, and the value function. The first two terms on the right-hand side are the usual terms from the static bidding model. They say that the bid equals the private cost plus a mark-up term that accounts for the level of competition in the current period. In the dynamic model, I have a second mark-up term that accounts for the incremental effect on the future discounted stream of payoffs that arises when firm i wins the contract instead of another firm. In other words, this term accounts for the inter-temporal trade-off faced by the firm: winning an auction today implies higher profits in the current period, but also an increase in the firms's backlog which leads to lower profits in the future (a higher backlog worsens the firm's strategic position and profits in future auctions).[38] These forgone future profits are taken into account by the firm when making its bidding decision.

In the pioneering approach of Guerre, Perrigne, and Vuong (2000) one would use the FOC to infer the distribution of private values. Take, for example, the FOC corresponding to the static model (first line of equation (10)). In that case bids are observed, and I have shown in the previous section that the conditional distribution function of equilibrium bids is identified. Therefore, all the objects on the right-hand

[37]This is in fact an abuse of language since the cost c_i actually contains some common knowledge components like Y and W.

[38]If instead we think that there is learning-by-doing then winning a project today would involve higher profits in the future.

side are identified which means that the conditional distribution of private values is also identified.[39] But with the FOC given in (10) we cannot proceed in this way since there is an extra term including the value function. The problem is that the value function given in equation (8) involves costs that are unobserved and equilibrium decisions which are endogenous.

In the following Lemma, I derive a result that allows me to express the value function as a recursive equation in terms of the distribution of equilibrium bids only. This is an extension of the result in JP without unobserved heterogeneity and exogenous firm participation. The proof is given in the Appendix.

Lemma 6.1

$$\begin{aligned} V_i(s) \;=\; & E_{YW\mathcal{N}}\Big\{1\{i\in\mathcal{N}\}\Big(\int \frac{1}{\sum_{j\in\mathcal{N},j\neq i} h(\cdot|y,w,\mathcal{N},s_j,s_{-j})} dG^{(i)}(\cdot|y,w,\mathcal{N},\boldsymbol{s}) \\ & +\beta\sum_{j\in\mathcal{N},j\neq i}\Big[Pr(j\ \ wins|y,w,\mathcal{N},s_i,s_{-i}) \\ & +\int\frac{h(\cdot|y,w,\mathcal{N},s_i,s_{-i})}{\sum_{l\in\mathcal{N},l\neq i} h(\cdot|y,w,\mathcal{N},s_l,s_{-l})} dG^{(j)}(\cdot|y,w,\mathcal{N},\boldsymbol{s})\Big] V_i(\omega(x,\boldsymbol{s},j))\Big) \\ & +1\{i\notin\mathcal{N}\}\beta\sum_{j\in\mathcal{N}} Pr(j\ \ wins|y,w,\mathcal{N},\boldsymbol{s})V_i(\omega(x,\boldsymbol{s},j))\Big\} \end{aligned} \tag{11}$$

where $G^{(i)}(b|y,w,\mathcal{N},s)$ is the ex-ante probability that i wins with a bid of b or less.

The representation given in Lemma 6.1 is a linear system of equations, with the following solution

$$V_i = [I-\beta(B_i+D_i)]^{-1}A_i \tag{12}$$

where A_i, B_i, and D_i are functions of the distribution of equilibrium bids and distributions of observable and unobservable characteristics. A_i is the vector of expected current period profits and (B_i+D_i) is the matrix of transition probabilities. See the Appendix for the derivation of (12) and expressions for A_i, B_i, and D_i.

From Theorem 6.2 and Lemma 6.1, all the objects on the right-hand size of the FOC are identified from the joint distribution of the observables. This implies that

[39] There is a difference between the approach in Guerre, Perrigne, and Vuong (2000) and the one I introduce here when I allow for unobserved heterogeneity. I can no longer recover the private cost c associated with every bid as they do, because I cannot recover the value of W for any particular auction. But this does not pose a problem, since I can still recover the distribution of private values conditional on any given value of W.

the distribution of private values is also identified. I summarize this result in Theorem 6.3.

Theorem 6.3 *Suppose Assumptions 5.1-5.4 and 6.1-6.3 hold. Suppose β is fixed and known to the econometrician. Then the conditional cost distribution functions, $F_{C|YWs}(\cdot|\cdot,\cdot)$ and $F^f_{C|YW}(\cdot|\cdot,\cdot)$ are nonparametrically identified.*

7 Estimation

In this section I describe the estimation procedure I take to the data. Although a nonparametric estimator is possible in principle, to avoid the curse of dimensionality I implement a parametric version. The estimator follows from the identification proof and it involves four steps. The first step estimates the control functions, i.e., I estimate the participation equations and recover estimates of w^A and w^B. The second step estimates the distribution functions of equilibrium bids conditional on observable characteristics, state variables and W^B. The third step implements the method from Hu (2008) using the estimates from the first two steps to estimate the distribution of equilibrium bids conditional on W. Finally, the last step estimates the value function and uses the FOC to recover the conditional distribution of private costs.

7.1 Estimation of the Control Functions

The estimation of the control functions follows HHS. I specify a parametric form for $\zeta_A(\cdot)$ and $\zeta_B(\cdot)$ and estimate the two participation equations via maximum likelihood. Here I describe how I estimate (5) and estimation of (3) follows the same procedure.

The model is nonstandard due to the discreteness and the truncation from below at $N^B = 1$. Suppose $\zeta^B(Y, \boldsymbol{s}) = \lfloor y\gamma_y^B + \boldsymbol{s}\gamma_s^B \rfloor$ where $\lfloor\cdot\rfloor$ is the floor operator.[40] No further assumption is made on W^B, which therefore has a multinomial distribution with probabilities $\boldsymbol{p^B} \equiv (p_1^B, \ldots, p_K^B)$ on $w_1^B, \ldots, w_K^B$. Imposing $p_K^B = 1 - \sum_{k=1}^{K-1} p_k^B$ and a median zero restriction on W^B, the likelihood function for the sample $\{(n_t^B, y_t, \boldsymbol{s}_t)\}_{t=1}^T$ is given by

$$(13) \quad \mathcal{L}(\boldsymbol{p}, \gamma^B) = \prod_{t=1}^{T} \frac{\sum_{i=1}^K p_i 1\{\lfloor y_t\gamma_y^B + \boldsymbol{s}_t\gamma_s^B \rfloor + w_i = n_t^B\}}{\sum_{i=1}^K p_i 1\{\underline{n}^B \leq \lfloor y_t\gamma_y^B + \boldsymbol{s}_t\gamma_s^B \rfloor\}}$$

[40]For $x \in \mathbb{R}$, $\lfloor x \rfloor = \max\{\iota \in \mathbb{Z} : \iota \leq x\}$.

subject to

$$(14)\quad w_1 \leq n_t - \lfloor y_t\gamma_y^B + \boldsymbol{s}_t\gamma_s^B \rfloor \leq w_K, \qquad \forall t.$$

The estimation is performed via maximum likelihood. The objective function is non-differentiable in some of the parameters and presents multiple local maxima. The computational details on how to overcome both issues in practice are given in Balat and Haile (2011).

Given the parameter estimates $\hat{\gamma}_y^A, \hat{\gamma}_s^A, \hat{\gamma}_y^B$ and $\hat{\gamma}_y^B$, I then get estimates of the control functions as follows:

$$\hat{w}_t^j = n_t^j - \lfloor y_t^j\hat{\gamma}_y^j + \boldsymbol{s}_t\hat{\gamma}_s^j \rfloor, \quad j = \{A, B\}$$

7.2 Distribution Function of Equilibrium Bids

An interim step requires the estimation of the bid distributions functions for regular and fringe bidders conditional on observed contract characteristics, state variables, and W_B. I follow JP, Athey, Levin, and Seira (2011) and Groeger (2010) by estimating these distributions parametrically under the assumption that they follow a log-Weibull distribution. The scale and shape parameters are both positive scalars and depend on observed contract characteristics, state variables, and W_B. Omitting the dependence on the covariates, the two density functions take the same general form

$$(15)\quad g^j(b|\theta^j) = \frac{1}{b}\left[\frac{\theta_1^j}{\theta_2^j}\left(\frac{\log(b)}{\theta_2^j}\right)^{\theta_1^j-1} e^{-\left(\frac{\log(b)}{\theta_2^j}\right)^{\theta_1^j}}\right], \qquad j = \{\text{regular}, \text{fringe}\}.$$

The specifications for the parameters θ are given in Section 8.

The estimation is performed via maximum likelihood, where the likelihood function takes the form

$$(16)\quad L = \prod_{t=1}^{T} \prod_{i\in\mathcal{N}_t, i\in\text{Reg}} g(b_{it}|\theta^R) \prod_{j\in\mathcal{N}_t, j\in\text{Fri}} g(b_{jt}|\theta^F).$$

7.3 Estimation of $g(b|Y, W, \mathcal{N}, \boldsymbol{s})$, $g^f(b|Y, W, \mathcal{N}, \boldsymbol{s})$ and $g(W)$

For notational simplicity I suppress the dependence on covariates and state variables. Here I show how I estimate the distribution of equilibrium bids for regular bidders conditional on W (the estimation for fringe bidders is analogous). The estimators in this section closely follow An, Hu, and Shum (2010) with two differences. First, the estimator for the conditional distribution of equilibrium bids is parametric in my case. Second, the two noisy measurements of W are estimated in the first stage instead of being observed in the data.

I begin by estimating the conditional distribution of W^B given W, $\mathbf{g_{W^B|W}}$, using the estimates from the previous two steps. My estimators for the matrices (33) and (35) in the Appendix are given by

$$(17) \quad \hat{\mathbf{g}}_{\mathbf{W^B},\mathbf{W^A}} = \left[\frac{1}{T} \sum_t 1\{\hat{w}_t^B = j, \hat{w}_t^A = k\} \right]_{j,k}$$

$$(18) \quad \hat{\mathbf{g}}_{\mathbf{Eb},\mathbf{W^B},\mathbf{W^A}} = \left[\hat{E}(b|W^B = j, W^A = k) \frac{1}{T} \sum_t 1\{\hat{w}_t^B = j, \hat{w}_t^A = k\} \right]_{j,k}$$

where $E(b|W^B = j, W^A = k)$ is estimated by numerical integration using an estimate of the distribution function of equilibrium bids conditional on both W^A and W^B, and evaluated at the mean values of the other covariates.

Then, I proceed to estimate $\mathbf{g_{W^B|W}}$ with

$$(19) \quad \hat{\mathbf{g}}_{\mathbf{W^B|W}} = \psi(\hat{\mathbf{g}}_{\mathbf{Eb},\mathbf{W^B},\mathbf{W^A}} \; \hat{\mathbf{g}}^{-1}_{\mathbf{W^B},\mathbf{W^A}})$$

where $\psi(\cdot)$ is a deterministic analytic function that takes the eigen-vector of the matrix $\hat{\mathbf{g}}_{\mathbf{Eb},\mathbf{W^B},\mathbf{W^A}} \; \hat{\mathbf{g}}^{-1}_{\mathbf{W^B},\mathbf{W^A}}$, normalizes each column to sum up to 1, and reorders them according to the order of the eigen-values given by (38) (which is implied by Assumption 5.1).

Finally, I estimate $g(W)$ and $g(b|W)$ pointwise in b. From equations (46), (47) and (48) in the Appendix, I propose the following estimators:

$$(20) \quad \hat{g}(W) = \hat{\mathbf{g}}^{-1}_{\mathbf{W^B|W}} \; \hat{\vec{g}}(W^B)$$

$$(21) \quad \hat{g}(b|W) = \frac{e'_W \; \hat{\mathbf{g}}^{-1}_{\mathbf{W^B|W}} \; \hat{\vec{g}}(b, W^B)}{e'_W \; \hat{\mathbf{g}}^{-1}_{\mathbf{W^B|W}} \; \hat{\vec{g}}(W^B)}$$

$$(22) \quad \hat{G}(b|W) = \frac{e'_W \; \hat{\mathbf{g}}^{-1}_{\mathbf{W^B|W}} \; \hat{\vec{G}}(b, W^B)}{e'_W \; \hat{\mathbf{g}}^{-1}_{\mathbf{W^B|W}} \; \hat{\vec{g}}(W^B)}$$

where

$$(23)\quad \hat{\vec{g}}(W^B) = \left[\frac{1}{T}\sum_t 1\{W_t^B = j\}\right]_j$$

$$(24)\quad \hat{\vec{g}}(b, W^B) = \left[\hat{g}(b|W^B = j)\hat{g}(W^B = j)\right]_j$$

$$(25)\quad \hat{\vec{G}}(b, W^B) = \left[\hat{G}(b|W^B = j)\hat{g}(W^B = j)\right]_j$$

7.4 Estimation of $F_{C|Y,W,s}$ and $F^f_{C|Y,W}$

Remember from the FOC at the bidding stage given in equation (10), that in order to estimate the distribution of private costs, we need estimates of the distribution and density functions of equilibrium bids and an estimate of the value function. To estimate the value function, I follow the procedure in JP, modifying it to include the new term D_i in (12). See the Appendix for details.

With an estimate of the value function at hand, I then fix W at some value (w_j) and draw a sample of bids from $\hat{G}(b|y, w_j, \mathcal{N}, s)$ and $\hat{G}^f(b|y, w_j, \mathcal{N})$ with the value of the other variables fixed the desired level (e.g., at their observed means or at the values of a specific project). Using that sample, I then use the FOC (10) to recover the private cost associated with each bid. Finally, by the monotonicity of the bidding function in c, I can write $\hat{F}_{C|YWs}(c|y, w_j, s_i) = \hat{G}_{b|YWs_is_{-i}}(\xi^{-1}(c|y, w_j, s_i, s_{-i})|y, w_j, s_i, s_{-i})$ and a similar expression for the distribution of fringe bidders.

8 Estimation Results

In this section I present the results from the multi-step estimator introduced in the previous section.

Reduced form participation equations. I start by showing the estimates of the reduced form equations (3) and (5). I implement the parametric specification as explained in Section 7.1 and the results are shown in Table 9. To allow for more flexibility and to improve the performance of the estimators, I use 20 points in the supports of W^A and W^B. However, when I take the estimates of W^A and W^B (i.e., estimates of the control functions) to the second step in the procedure, I reduce the dimensionality by grouping contiguous values so that both control functions have 4 points in their support in the second stage. Note that Assumption 5.1 still holds at the group level.

Table 9: Reduced Form Participation Equations

	N^A	N^B
constant	8.023	3.201
log(eng)	-.245	-.034
log(days)	.407	-.019
items	.011	-.0001
costindex	-.010	-.0001
pot bidders	.012	.006
var cty	-.572	-.310
sum std bl	-.189	-.102
nobs	5164	5164

All regressions include time, district, and type of work dummies.

The results show that larger projects attract fewer planholders and result in fewer number of bidders. While one may expect a positive sign a priori, the result is consistent with the descriptive regressions in Table 7. A plausible explanation is that while regular firms do participate more in larger contracts (see Table 8), fringe firms do not (see Table 17 in the Appendix). Similar results are found by Krasnokutskaya and Seim (2011). Regarding the effect of backlog, I find as expected that the higher the level of backlogs, the fewer the number of planholders and bidders.

Bid density estimates. In order to estimate the conditional distribution and corresponding density functions of equilibrium bids for regular and fringe firms, several steps are required. The estimation results for the interim steps are in the Appendix. One key step, though, is the bid density estimations conditional on the state variables, observable project characteristics, and W^B, a misclassified measurement of W. I model the scale and shape parameters of the Weibull distributions, denoted θ_1^j and θ_2^j, respectively, for $j = \{\text{regular}, \text{fringe}\}$ as follows. The parameters of the density function for a regular bidder are given by

$$(26) \quad \log(\theta_1^R) = \gamma_{1,0}^R + \sum_{k=1}^{K-1} \gamma_{1,k}^R 1\{\hat{w}_b = k\}$$

$$
\begin{aligned}
\log(\theta_2^R) \quad = \quad & \gamma_{2,0}^R + \textstyle\sum_{k=1}^{K-1} \gamma_{2,k}^R 1\{\hat{w}_b = k\} + \gamma_{2,K}^R \log(\text{estimate}) \\
& + \gamma_{2,K+1}^R \log(\text{days}) + \gamma_{2,K+2}^R \log(\text{dist}) \\
& + \gamma_{2,K+3}^R \#\text{fringe} + \gamma_{2,K+4}^R \#\text{regular} \\
& + \gamma_{2,K+5}^R \text{stdbl} + \gamma_{2,K+6}^R \text{sumstdblin} + \gamma_{2,K+7}^R \text{sumstdblout}
\end{aligned} \tag{27}
$$

where stdbl denotes (regular) firm i's standardized backlog, sumstdblin $= \sum_{j \neq i, j \in \mathcal{N}_t} \text{stdbl}_j$ (i.e., the sum of the standardized backlogs for rival regular firms in the auction) and sumstdblout $= \sum_{j \notin \mathcal{N}_t} \text{stdbl}_j$ (i.e., the sum of the standardized backlogs for regular firms not participating in the auction). The rest of the variables are defined as before.

Similarly, the density function for fringe bidders has parameters given by

$$
\log(\theta_1^F) = \gamma_{1,0}^F + \sum_{k=1}^{K-1} \gamma_{1,k}^F 1\{\hat{w}_b = k\} \tag{28}
$$

$$
\begin{aligned}
\log(\theta_2^F) \quad = \quad & \gamma_{2,0}^F + \textstyle\sum_{k=1}^{K-1} \gamma_{2,k}^F 1\{\hat{w}_b = k\} + \gamma_{2,K}^F \log(\text{estimate}) \\
& + \gamma_{2,K+1}^F \log(\text{days}) + \gamma_{2,K+2}^F \log(\text{dist}) \\
& + \gamma_{2,K+3}^F \#\text{fringe} + \gamma_{2,K+4}^F \#\text{regular} \\
& + \gamma_{2,K+5}^F \text{sumstdblin} + \gamma_{2,K+6}^F \text{sumstdblout}
\end{aligned} \tag{29}
$$

where now sumstdblin $= \sum_{j \in \mathcal{N}_t} \text{stdbl}_j$.

Additionally, the two scale parameters, θ_2^R and θ_2^F, include district and year dummies. I include the sum of the backlogs and not the individual components because the model imposes a symmetry condition on the bidding function conditional on the state variables. Bidders with the same state follow the same bidding strategy. Hence, the order of the elements of the vector of backlogs should not affect the parameters of the regular and fringe bidders' distribution. This implies that the coefficients for the backlogs should be the same. However, I distinguish between regular firms participating in the auction and those who are not, and thus I allow for different coefficients for the two groups.

Table 10 shows the parameter estimates. All signs are as expected. Note that since W^B is correlated with W it should partially control for the unobserved heterogeneity, but not fully. Nevertheless, when I condition on W using the estimator in Section 7.3 I get the same qualitative results. Note that θ_2 is the scale parameter for the Weibull distributions, thus a positive sign means that the distribution is stretched to the right, and a negative that it is shrunk to the left. In other words, a positive coefficient indicates that the distribution associated with a higher value of the variable in question stochastically dominates (in the first-order sense) the distribution associated with a lower value of the variable. The length of the project and the distance from the

Table 10: Bid Density Estimates

	θ_1^R	θ_2^R	θ_1^F	θ_2^F
constant	-2.1206e+00	-1.4539e+00	-1.5195e+00	-1.7622e+00
log(days)		1.0340e-03		1.5216e-03
log(eng)	3.3070e-01	2.1079e-01	2.8545e-01	2.2787e-01
log(dist)		5.8352e-05		3.7980e-05
# fringe		-1.2521e-02		-9.3289e-03
# regular		-1.2529e-02		-9.5619e-03
std bl		4.4770e-03		
sum std bl $\in \mathcal{N}$		1.9336e-03		4.1775e-03
sum std bl $\notin \mathcal{N}$		1.1832e-03		1.0520e-03
$W^B = 1$	1.6640e-01	-7.3694e-02	1.0149e-02	-2.3398e-02
$W^B = 2$	1.5341e-01	-5.7130e-02	7.2210e-02	-2.2929e-02
$W^B = 3$	1.7491e-01	-2.6712e-02	1.0750e-01	-1.0199e-02

All specifications include time and district dummies.

firm to the work site shift both distributions of equilibrium bids to the right. We also see that there is a competition effect, in the sense that a higher number of bidders shift the distributions to the left. More importantly, a firm's backlog and their rival's backlogs enter with a positive sign. This means that a regular firm is more likely to submit a larger bid the larger its backlog, and also that both regular and fringe firms are more likely to submit a larger bid when their rivals have larger backlogs.

Unobserved heterogeneity. I plot the distribution function of equilibrium bids for a regular bidder conditional on the true unobserved factor (evaluated at the mean values of the other variables) in Figure 2. Figure 3 plots the corresponding densities. Controlling for unobserved heterogeneity significantly affects the distribution of bids. As W increases, the distribution stretches to the right but also changes in shape. The effects appear to be nonlinear. It also appears that the dispersion of bids increases with W. One plausible explanation is as follows. If we think of the unobserved heterogeneity as the quality of the existing road to be repaired, a road in a very bad condition (a high W) will translate in higher repair costs and thus a higher bid. But it may also introduce a greater dispersion in the repair costs, since bad quality roads could require nonstandard methods to fix them and firms can differ in their expertise or experience for fixing bad quality roads. Good quality roads may only require standard procedures, thus the variance in firms' ability to fix them may be smaller.

Table 11 shows the estimated distribution of the unobserved heterogeneity W. It also shows the expected bid conditional on W (evaluated at the mean of other

Figure 2: Distribution Function of Equilibrium Bids for a Regular Bidder

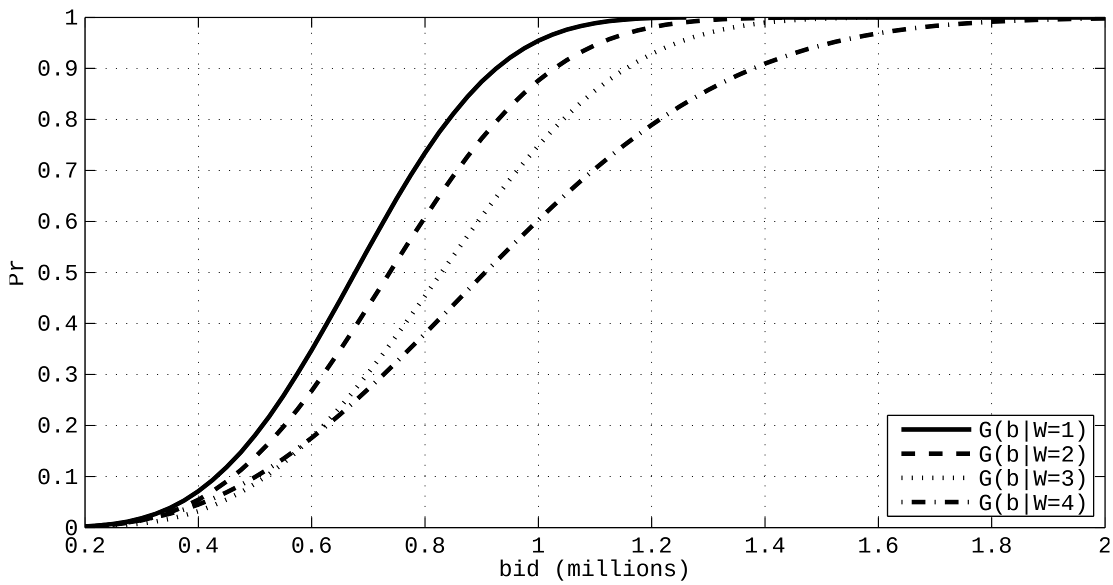


Figure 3: Density Function of Equilibrium Bids for a Regular Bidder

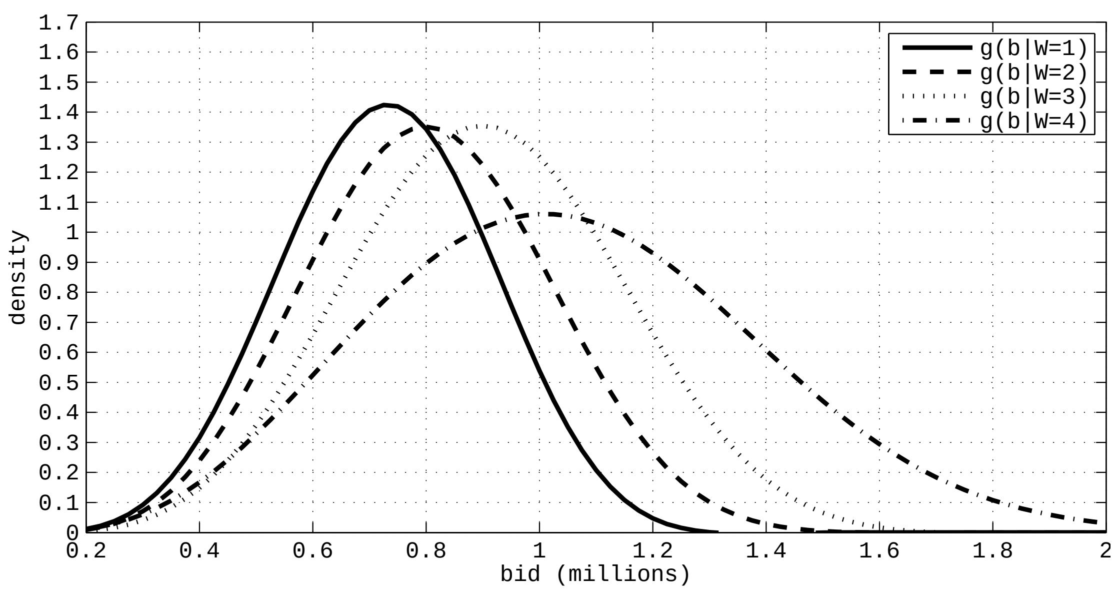


Table 11: Unobserved Heterogeneity

W	$\hat{\Pr}(W)$	$\hat{E}[B\|W]$	$\hat{E}[B^1\|W]$ (in millions)	$\hat{E}[C\|W]$	$\frac{(b-c)}{b}$
1	.0762	.7305	.3929	.6536	.1446
2	.3516	.8069	.4181	.7440	.1140
3	.3226	.9047	.4450	.8510	.0891
4	.2496	1.0438	.4524	.9911	.0847

variables) in the second column. The expected equilibrium bid for a regular bidder varies with W in a monotonic but nonlinear way. Changing W from a value of 1 to 2 produces an increase in the expected bid of 8.6%, a change from 2 to 3, an increase of 13%, and lastly, a change from 3 to 4, is associated with an increase of 9.3%. Column 3 shows the expected winning (i.e., lowest) bid, column 4 the expected cost. The expected cost for a regular firm and the equilibrium price in the auction also increase with W. In particular, the cost when the unobserved heterogeneity is at its highest level is 52% higher than when the unobserved heterogeneity is at its lowest level. The equilibrium price increases by 15% when W goes from 1 to 4.

Distribution of private costs. From the last step in the estimation procedure, I recover the conditional distributions of private costs. I show the distribution for a regular bidder in Figure 4. In some cases,[41] the inferred cost is negative which is not possible. In such a situation, I set the cost equal to zero. The same remarks made for the distributions of equilibrium bids also apply for the distribution of private costs.

Bid Function and Mark-up. Figure 5 shows the bid functions estimated using the FOC for one of the regular bidders. The bid function is plotted by fixing the state variables at sample average values and varying the cost. The figure also plots the 45 degree line. Conditional on W, the distance between the bid and the cost decreases as we increase the cost, but contrary to the usual result in static models, the bid does not approach the 45 degree line, a result also found by JP. This can be attributed to the mark-up term involving the value function in (10), i.e., the negative effect on the future discounted profits if firm i wins the contract instead of another firm. Note that while the bid functions are increasing in the unobserved heterogeneity, the average mark-up is decreasing in it (see the last column of Table 11). The reason is that higher values of W are associated with higher realizations of costs, and the mark-ups are decreasing in cost as discussed above.

[41] It depends on the characteristics of the projects, but on average around 5% of the cost predictions are negative.

Figure 4: Distribution Function of Private Cost for a Regular Bidder

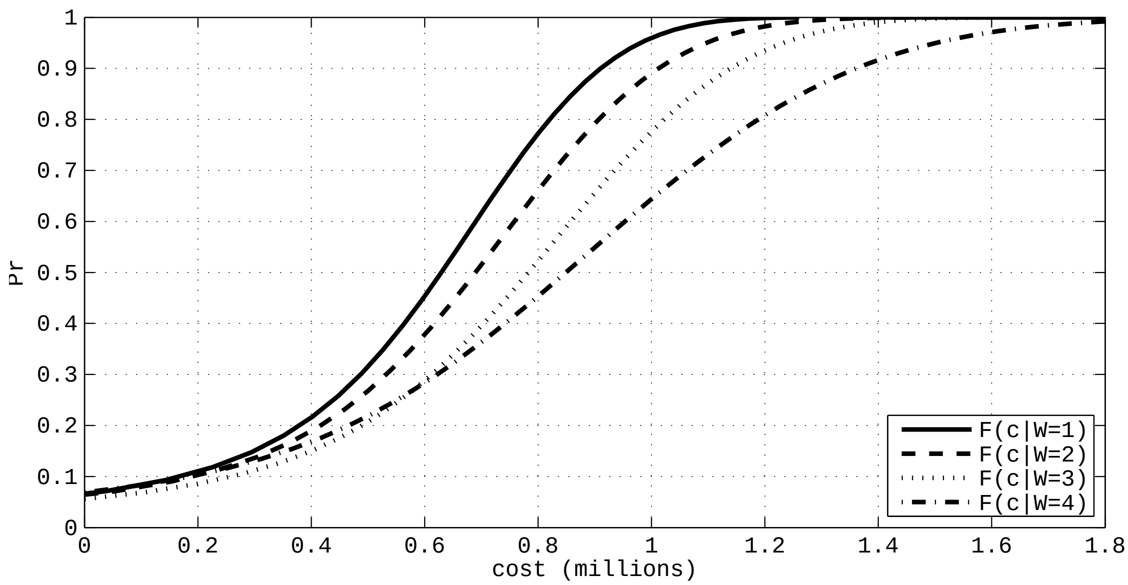


Figure 5: Bid Function for a Regular Bidder

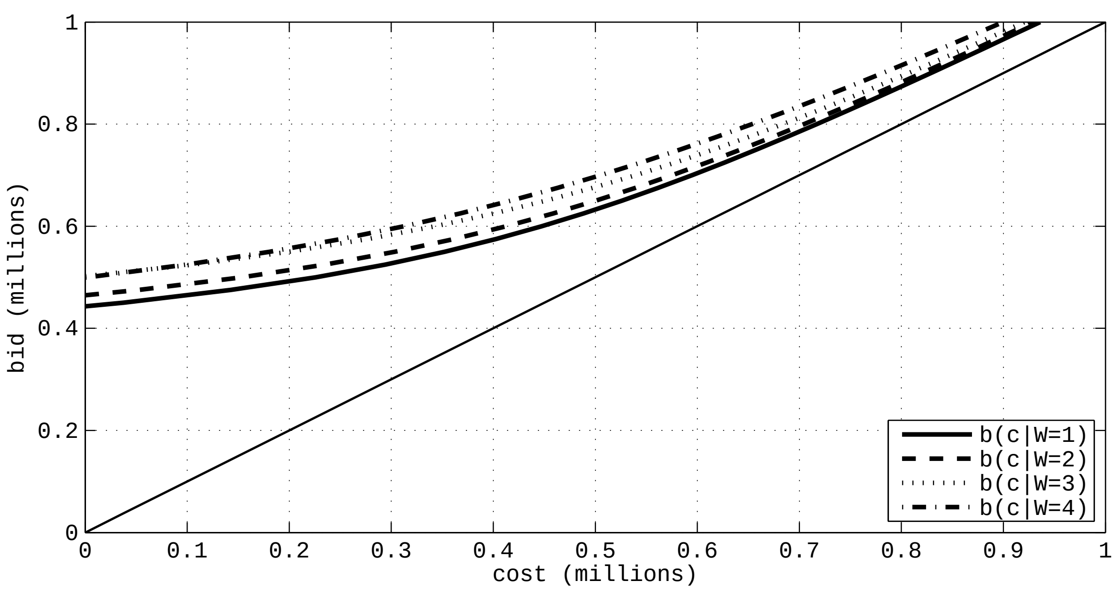


Table 12: The Effect of Backlog

W	backlog$_i = -1$ $\hat{E}[B\|W]$	$\hat{E}[B^1\|W]$ (in millions)	$\hat{E}[C\|W]$	$\frac{(b-c)}{b}$	backlog$_i = +1$ $\hat{E}[B\|W]$	$\hat{E}[B^1\|W]$ (in millions)	$\hat{E}[C\|W]$	$\frac{(b-c)}{b}$
1	0.7061	0.3819	0.6293	0.1488	0.7531	0.4023	0.6753	0.1428
2	0.7768	0.4014	0.7133	0.1185	0.8301	0.4256	0.7663	0.1128
3	0.8700	0.4379	0.8157	0.0927	0.9315	0.4552	0.8771	0.0878
4	1.0041	0.4391	0.9510	0.0883	1.0746	0.4570	1.0211	0.0838

Effects of Backlog. To asses the effect of the backlog, I perform the following two exercises. First, I take one regular firm and compute its expected bid and cost under the assumption that the firm has a backlog 1 standard deviation above its mean. Then I perform the same calculations but assume the firm has a backlog equal to 1 standard deviation below its mean. All project characteristics are kept at their observed means (including backlogs of other firms), and I assume that there are 2 regular firms participating in the auction and 7 fringe firms. I also calculate the auction equilibrium price in this typical auction, under the two backlog levels for the regular firm. The effect on the bid of this regular bidder whose backlog I change only includes the direct channel mentioned earlier, since the backlogs of other firms are not changing. On the other hand, the auction equilibrium price includes all three channels, the direct effect, the strategic effect and the competition effect[42] since I allow the other firms to respond to the change in the backlog of the regular firm.

Results are presented in Table 12. As expected, conditional on W, both the mean bid and cost are higher when the firm has a higher backlog, but the mean mark-up decreases, indicating that costs increase more than bids. Averaging over W, the bid increases by 6.9%, and the cost by 7.4%. The equilibrium price in the auction increases by 4.7% on average.

The second exercise I perform considers a change in the backlogs of all 10 regular firms from 1 standard deviation below the mean to 1 standard deviation above the mean. The rest of the setup is similar to the one in the previous exercise. Results are presented in Table 13. The average bid for the same regular firm as in the previous exercise increases now by 26.8%. This change now includes the strategic and competition effects, which in this case have sizable impacts. The equilibrium

[42]Since the change in backlog we are considering is for only one firm, it is not enough in practice to alter the number of firms participating in the auction. Thus the competition effect in this exercise is zero.

Table 13: The Effect of Backlog

W	all backlogs = −1				all backlogs = +1			
	$\hat{E}[B\|W]$	$\hat{E}[B^1\|W]$ (in millions)	$\hat{E}[C\|W]$	$\frac{(b-c)}{b}$	$\hat{E}[B\|W]$	$\hat{E}[B^1\|W]$ (in millions)	$\hat{E}[C\|W]$	$\frac{(b-c)}{b}$
1	0.6403	0.3577	0.6049	0.0920	0.8036	0.4275	0.6929	0.1772
2	0.7025	0.3794	0.6890	0.0526	0.8878	0.4478	0.7983	0.1370
3	0.7850	0.3997	0.7918	0.0174	0.9993	0.4899	0.9283	0.1010
4	0.9023	0.3999	0.9219	0.0068	1.1456	0.4901	1.0858	0.0857

price in the auction increases by 20.8%.

9 Simulation Results

9.1 The effect of the stimulus package on procurement costs

The key insight of this paper is that the injection of funds into the economy in a short period of time affects firms' backlogs, which in turn drives up firms' costs and prices paid by the government. In this section, I analyze how much higher were procurement costs for stimulus funded projects and for projects funded from other sources due to the effect of stimulus projects on firms' backlogs.

To answer this question, I use the estimates of the primitives of my structural model in a counterfactual simulation to eliminate the effect of stimulus projects on backlogs. I then compare the equilibrium project prices under the counterfactual to those of the baseline case in which the stimulus projects do affect firms' backlogs. Details of how the simulations are performed are presented in the Appendix. The results are presented in Table 14.

I find that the government paid prices for stimulus funded projects that are, on average, 6.2% higher than in the counterfactual case. This represents $105.4 million worth of extra projects that the government could have acquired by facing the lower prices. The effect on prices for other projects not funded by the stimulus package is sizable. On average, prices for these projects were 4.8% higher than in the counterfactual, which implies $229.3 million worth of projects the government has forgone due to the backlog effect of the stimulus projects.[43]

[43]These results are in line with the theoretical findings of Saini (2011). He finds that the procurer faces lower procurement costs by scheduling frequent auctions for small project sizes.

Table 14: Effect of the Stimulus on Prices

Projects	Δp	$\Delta\$$ (M)
ARRA ($1.7B)	6.2%	105.4
Other ($4.8B)	4.8%	229.3
Total		334.7

The way to interpret these results is as follows. Instead of injecting the funds *all-at-once*[44] the government could have "spread out" the expenditure as much as needed, potentially infinitely long, so that every project could have been completed within a period, thus having no impact on firms' backlogs. Therefore, one can consider the total dollar value of forgone projects, $334.7 million, as the "cost" of ARRA's stimulus effect. The total toll on forgone projects from releasing the stimulus projects at an accelerated pace represents 19.7% of the stimulus funds received by California. By no means does this imply that $334.7 million were "lost," since the money was actually transferred to the firms, thus serving the government's stimulus objective. The result just makes explicit that part of the demand expansion went into higher prices rather than quantities.

9.2 The effect of the stimulus package on the cost of production

The actual cost of completing a project (i.e., the cost of the winning firm) and the procurement price paid by the government are not the same because of the mark-up terms involving strategic effects and the option value (see Section 6.3). In this section I quantify how much higher were firms' costs due to the effect of the stimulus projects on backlogs.

From a welfare perspective, if one were to ignore the cost of raising the money to finance the stimulus package, one would only care about the cost of the resources used to build or repair the roads. Due to the effect of backlogs on firms' costs, we should expect an inefficiency generated by the stimulus funds.

I perform the same simulations as in the previous section and calculate the effect of the stimulus projects on the winning firm's cost rather than its bid. In terms of the cost of the resources used, the baseline case involves costs that are, on average, 2.8%

[44]From the discussion in Section 3, the way the projects were released in practice comes close to the notion of spending all the money at once, given the administrative and physical constraints.

Table 15: Effect of the Stimulus on Prices (Lower Backlogs)

Projects	Δp	$\Delta\$$ (M)
ARRA ($1.7B)	5.68%	96.5
Other ($4.8B)	4.27%	203.9
Total		300.4

Backlogs of all regular firms are set at 1 standard deviation below their historical average at the time the stimulus projects were first released.

higher than in the counterfactual. This implies that the total cost of projects (from all funding sources) increased by $151 million, an amount that represents 8.8% of the stimulus funds received by California. Also this exercise allows us to decompose the effect on procurement prices in its two components: the change in firms' costs and the change in mark-ups. I find that, on average, 45% of the change in price is due to higher costs and the remaining 55% due to higher mark-ups.

9.3 Was California special?

Here I examine whether there was something special about California that might drive the results in the previous simulations. This is particularly important if one wants to extrapolate the results from this paper to other applications.

In Table 1 we do not observe a downturn in the highway procurement sector as a result of the crisis during 2008. In the case of California, the absence of a downturn is explained by the fact that in 2006, well before the crisis started, a $20 billion transportation bond (called Proposition 1B) was approved by voters to support the funding of highway and road repair projects. As a result, when the stimulus projects were released, backlog levels were comparable to pre-crisis levels. This is not the typical scenario we expect to see when implementing a stimulus demand expansion.

Backlog levels. First, I explore the extent to which the "high" backlogs are responsible for the results in the previous section. In particular, I ask what happens to the previous results had California experienced a downturn in the highway sector by the end of 2008. I rerun the previous simulations assuming the backlogs of all regular firms were arbitrarily set at 1 standard deviation below their historical average at the time the stimulus projects were first released.

Table 15 shows the results. I find that stimulus project prices in the new baseline scenario —although lower than the prices observed in the previous baseline case due to the lower backlogs— are still 5.68% higher than under the new counterfactual. Prices for other projects in the baseline case are 4.27% higher than under the counterfactual. The total forgone projects due to the higher prices totals $300.5 million, or 17.67% of the stimulus funds.

This is not surprising since the effect of backlogs on procurement prices looks almost linear. In Figure 6 I plot the price of a project (with characteristics set at the observed means) in an auction with one regular bidder and 7 fringe bidders as a function of the (standardized) backlog of the regular firm. I keep the backlogs of other regular firms fixed. Even if the aggregate level of backlogs changes, I still get a close to linear relationship. Figure 7 shows the price of the same project but in an auction with 2 regular firms and 6 fringe bidders. I assume that all regular firms have the same backlog level, thus while varying the backlog of one firm I also vary the aggregate backlog level.

Funding levels. The second check I perform on my results is on the budget level for the highway construction sector. Most of the effect of the stimulus projects, as measured in terms of forgone projects, comes from higher prices for projects funded by other sources. Again, Proposition 1B may have kept the funding for those projects high throughout the crisis.

I analyze what happens to the results in the previous section had California experienced a higher cut in its budget. I take the case of Texas as a reference point, since Texas received stimulus funds comparable to California, but total expenditure on highway and road construction suffered a hefty cut during 2009 and 2010 (compared to total expenditure in 2008, there was a 18% cut in 2009 and a 35% cut in 2010).

In a back-of-the-envelope calculation, I use the prices changes from Section 9.1, but change the total budget level in 2009 and 2010. In particular, fixing the stimulus funds at the observed level, but adjusting the funds from other sources so that the total level of expenditure matches the cuts observed in Texas, I find that the effect of the price change for other projects now represents $169 million, and the total effect considering both stimulus projects and other projects accounts for $274.4 million, or 16% of the stimulus funds.

Figure 6: The Effect of Backlog on Procurement Prices

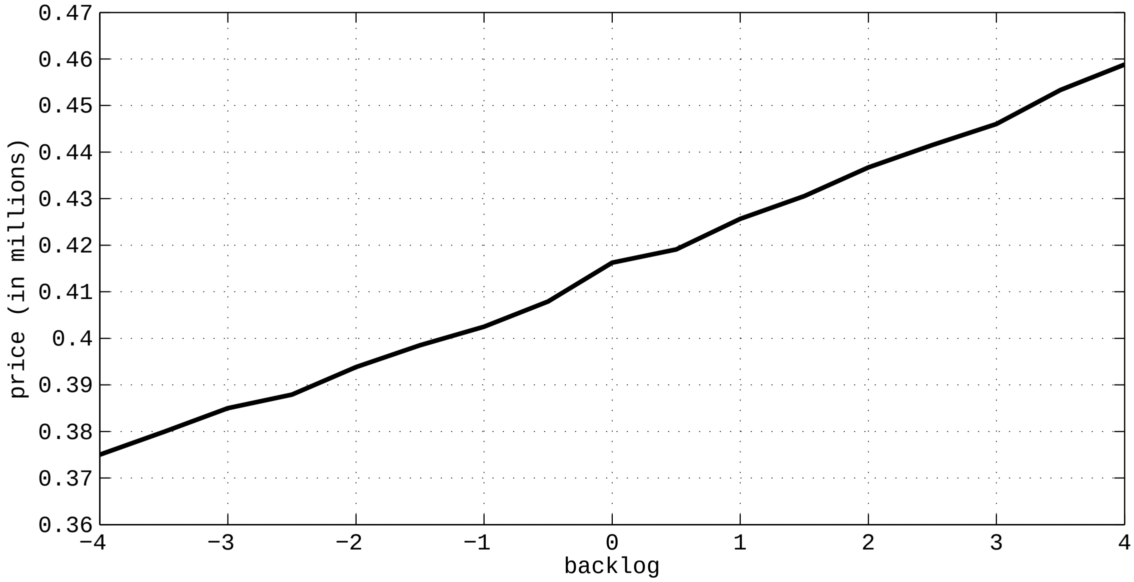
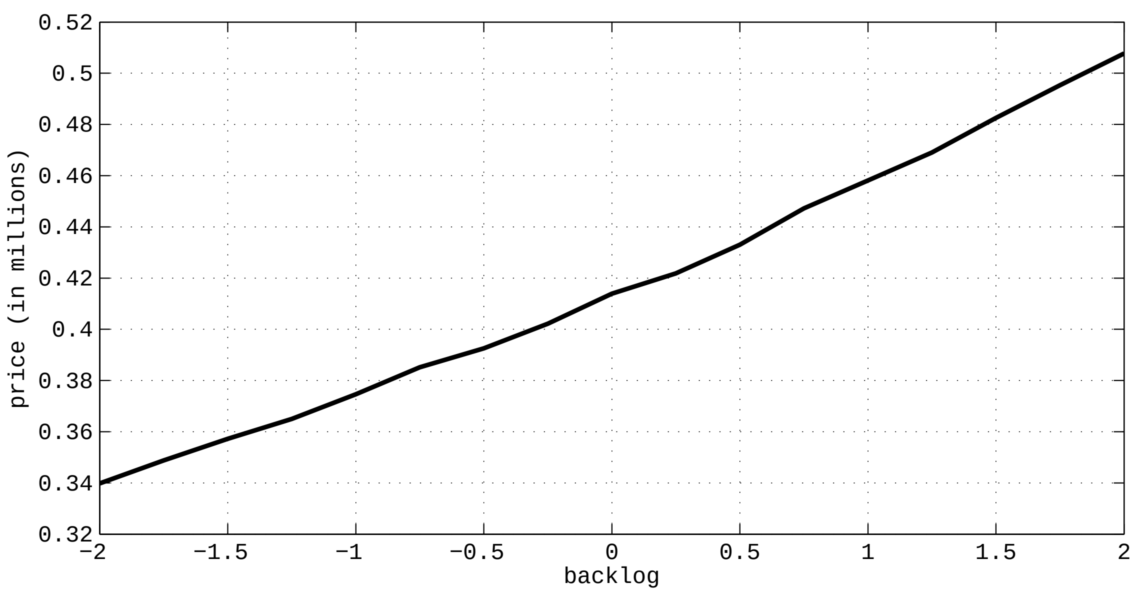


Figure 7: The Effect of Aggregate Backlog on Procurement Prices



9.4 The effect of a short-term delay in the stimulus on procurement costs

In this section I examine what the effect is on procurement prices from delaying the release of the stimulus projects by 3 and 6 months, while keeping the pace at which they were released unchanged.

While the results in Section 9.1 provide a natural way to describe how much prices were pushed up by injecting funds at an accelerated pace, the counterfactual neglects any stimulus effect at all. In this section I consider what happens with project prices by forgoing the stimulus effect only in the short-run. Details on the simulations are in the Appendix.

The simulation results for the effects of the stimulus package show that the government paid prices for stimulus funded projects that are on average 1.1% higher compared to those that were delayed 3 months, and 1.7% higher for those delayed 6 months. This represents, respectively, $18.7 and $28.9 million worth of projects the government had to give up to avoid a delay in the stimulus objective. Prices of other projects are 0.5% higher, on average, compared to those that result from delaying the ARRA projects by 3 months, and 0.6% higher for 6-month delay. This accounts for extra expenditure of $23.9 million and $57.5 million, respectively.

The total amount of public goods the government gave up (in dollar value) reaches $42.6 for the 3-month delay, and $57.5 million the 6-moth delay alternative. These values represent 2.5% and 3.4% of the stimulus funds, respectively. Even though no government would consider neglecting the stimulus effect, especially in the short-run, the results provide a measure of the opportunity cost of not delaying the projects. The rationale behind the counterfactuals is that by delaying the start of the stimulus projects the government lets the backlogs "die off" as previous committed work progresses, and hence at the time of starting the stimulus projects, faces a lower level of backlogs.

10 Conclusions

This paper considers the effects of the stimulus package on equilibrium prices paid by the government for highway construction projects. Using data from California, I answer (1) How much were the costs of these projects driven up by the accelerated pace of new projects? (2) What was the effect of the demand expansion on the prices of state funded projects that came afterwards? and (3) What was the effect on efficiency?. To this end, I develop a structural dynamic auction model that builds on three key features: upward sloping marginal costs, auction-level unobserved heterogeneity, and endogenous participation.

I show that the structural model is nonparametrically identified using concepts from the control function and measurement error literatures. I use the first order condition at the bidding stage to express each firm's private cost as a function of its bid, the conditional distribution of equilibrium bids, and the value function representing the discounted sum of future payoffs. The proof combines several key ideas. First, I show that the value function can be written as a function of the distribution of equilibrium bids. A second key idea is similar in spirit to the control function approach, but I allow for an "imperfect" control function. I exploit features of the procurement setting that provide a second "imperfect" control function. The information obtained from these two noisy controls then resembles a measurement error problem where we have access to multiple measurements. I attain identification of the conditional distribution functions using the results in Hu (2008) for nonlinear models with misclassification error.

From a methodological point of view, this paper contributes to the auction literature in several ways. I improve on the method of JP by controlling for unobserved heterogeneity, and by relaxing the assumptions on firms' participation decisions allowing for endogenous participation. To my knowledge, this is the first attempt to control for unobserved heterogeneity in a dynamic auction model. I also relax the structural assumptions in the control function approaches used previously in auction settings. In my model I allow the unobserved heterogeneity to enter nonlinearly in the firm's cost and I let the idiosyncratic component of the firm's cost to be correlated with the unobserved component.

From a policy perspective, this paper contributes to the discussion about the stimulus package by raising questions that have not been addressed yet. I quantify costs associated with the stimulus projects that may help in policy decisions. Counterfactual simulations indicate that the government has paid prices for stimulus funded projects that were 6.2% higher (and prices for other projects that were 4.8% higher) due to the effect of the stimulus projects on firms backlogs. These results imply that the government could have acquired $335 million worth of extra road projects (or 19.7% of the stimulus money received by California) by forgoing any stimulus effect from ARRA. Furthermore, I also show that even though California presents features that may distinguish it from other states, the results are robust and can be extrapolated to other applications. In a separate set of simulations I find that the opportunity cost of delaying the stimulus projects by 3 months reaches \$44 million (or 2.6% of the stimulus funds). If the government delays the projects by 6 months instead, the opportunity cost totals \$62 million (or 3.7% of the stimulus funds).

Appendices

A Proofs

A.1 Theorem 6.2

Proof Two remarks: (i) the proof holds conditional on observable characteristics and state variables, so I omit the conditioning variables for simplicity, and (ii) for ease of exposition, and without loss of generality, I assume $K^B = K$. The identification proof still holds for $K < K^B$ but I need to replace the matrix inverse operator with generalized inverses for non-square matrices.

Let me first introduce some notation. Let $g(\cdot)$ generally denote a probability mass or density function. Define the following K dimensional square matrices:

$$\mathbf{g_{b,W^B,W^A}} \equiv [g(b, W^B = i, W^A = j)]_{i,j} \tag{30}$$

$$\mathbf{g_{W^B|W}} \equiv [g(W^B = i|W = k)]_{i,k} \tag{31}$$

$$\mathbf{g_{W,W^A}} \equiv [g(W = k, W^A = j)]_{k,j} \tag{32}$$

$$\mathbf{g_{W^B,W^A}} \equiv [g(W^B = i, W^A = j)]_{i,j} \tag{33}$$

$$\mathbf{g_{b|W}} \equiv \begin{pmatrix} g(b|W=1) & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & g(b|W=K) \end{pmatrix} \tag{34}$$

$$\mathbf{g_{Eb,W^B,W^A}} \equiv [E(b|W^B = i, W^A = j)g(W^B = j, W^A = j)]_{i,j} \tag{35}$$

$$\mathbf{g_{Eb|W}} \equiv \begin{pmatrix} E(b|W=1) & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & E(b|W=K) \end{pmatrix} \tag{36}$$

Under assumptions 5.2-5.4 and 6.1 $\mathbf{g_{W^B,W^A}}$ is identified by Theorem 6.1. By a similar argument, $\mathbf{g_{b,W^B,W^A}}$ and $\mathbf{g_{Eb,W^B,W^A}}$ are identified since bids are observable.

The rest of the proof closely follows Hu (2008). It is constructive and provides a basis for the estimator developed in Section 7.

First, notice that

$$\begin{aligned} f(c|W, W^A, W^B) &= \frac{f(c, W^A, W^B|W)}{f(W^A, W^B|W)} \\ &= \frac{f(c|W)f(W^A|W)f(W^B|W)}{f(W^A|W)f(W^B|W)} \\ &= f(c|W) \end{aligned}$$

where the second equality follows from Assumption 6.2, and

$$\begin{aligned} g(W^B|W,W^A) &= \frac{g(W^B,W^A|W)}{g(W^A|W)} \\ &= \frac{g(W^B|W)g(W^A|W)}{g(W^A|W)} \\ &= g(W^B|W) \end{aligned}$$

again by Assumption 6.2.

By monotonicity of the bidding function in c, it is straightforward that

(37) $g(b|W,W^A,W^B) = g(b|W)$

and that

(38) $E[b|W=w_1] < E[b|W=w_2] < \ldots < E[b|W=w_K]$.

Next, I show identification of $\mathbf{g_{W^B|W}}$. Note that we can write

$$\begin{aligned} E[b|W^B,W^A]g(W^B,W^A) &= \textstyle\sum_{w=1}^K E[b|w,W^B,W^A]g(W^B,W^A|w)g(w) \\ &= \textstyle\sum_{w=1}^K E[b|w]g(W^B|w)g(w,W^A) \end{aligned}$$

where in the second equality I used (37), and we can also write

$$\begin{aligned} g(W^B,W^A) &= \textstyle\sum_{w=1}^K g(W^B|w,W^A)g(w,W^A) \\ &= \textstyle\sum_{w=1}^K g(W^B|w)g(w,W^A) \end{aligned}$$

using (38). In matrix notation,

(39) $\mathbf{g_{Eb,W^B,W^A}} = \mathbf{g_{W^B|W}}\ \mathbf{g_{Eb|W}}\ \mathbf{g_{W,W^A}}$

(40) $\mathbf{g_{W^B,W^A}} = \mathbf{g_{W^B|W}}\ \mathbf{g_{W,W^A}}$.

From (40) it follows that

$$\mathrm{Rank}(\mathbf{g_{W^B,W^A}}) \leq \min\{\mathrm{Rank}(\mathbf{g_{W^B|W}}), \mathrm{Rank}(\mathbf{g_{W,W^A}})\}$$

hence by Assumption 6.3, $\mathbf{g_{W^B|W}}$ and $\mathbf{g_{W,W^A}}$ are full rank. Therefore, I can invert (40) and write

(41) $\mathbf{g^{-1}_{W^B,W^A}} = \mathbf{g^{-1}_{W,W^A}}\ \mathbf{g^{-1}_{W^B|W}}$.

Postmultiplying (39) by (41) I get

(42) $\mathbf{g_{Eb,W^B,W^A}}\ \mathbf{g^{-1}_{W^B,W^A}} = \mathbf{g_{W^B|W}}\ \mathbf{g_{Eb|W}}\ \mathbf{g^{-1}_{W^B|W}}$.

Note that the last equation is an eigenvalue-eigenvector decomposition of the identified matrix $\mathbf{g_{Eb,W^B,W^A}}\ \mathbf{g^{-1}_{W^B,W^A}}$. This decomposition exists and is unique only if all the eigenvalues are distinct, which is guaranteed by Assumption 5.1.[45] Hence, $\mathbf{g_{Eb|W}}$ and $\mathbf{g_{W^B|W}}$ are identified.

Let $e_W = (0, \ldots, 0, 1, 0, \ldots, 0)'$, where the 1 is in the W-th entry of the vector, and define

$$\vec{g}(b, W^B) \equiv [g(b, W^B = 1), \ldots, g(b, W^B = K)]' \tag{43}$$

$$\vec{g}(W^B) \equiv [\Pr(W^B = 1), \ldots, \Pr(W^B = K)]'. \tag{44}$$

Then

$$g(b, W) = e'_W\ \mathbf{g^{-1}_{W^B|W}}\ \vec{g}(b, W^B) \tag{45}$$

and

$$g(W) = e'_W\ \mathbf{g^{-1}_{W^B|W}}\ \vec{g}(W^B) \tag{46}$$

hence $g(W)$ and $g(b, W)$ are identified.

Finally, identification of $g(b|W)$ and $G(b|W)$ is straightforward since

$$g(b|W) = \frac{e'_W\ \mathbf{g^{-1}_{W^B|W}}\ \vec{g}(b, W^B)}{e'_W\ \mathbf{g^{-1}_{W^B|W}}\ \vec{g}(W^B)} \tag{47}$$

$$G(b|W) = \frac{e'_W\ \mathbf{g^{-1}_{W^B|W}}\ \vec{G}(b, W^B)}{e'_W\ \mathbf{g^{-1}_{W^B|W}}\ \vec{g}(W^B)} \tag{48}$$

$\square$

A.2 Derivation of equation (12)

Recall that the state space is finite. Let m be the number of points in the support of state vector s, and denote each point with a superindex: $\boldsymbol{s}^1, \ldots, \boldsymbol{s}^m$. Define

$$A_i(\boldsymbol{s}) = E_{YW\mathcal{N}} 1\{i \in \mathcal{N}\} \int \frac{1}{\sum_{j\in\mathcal{N}, j\neq i} h(\cdot|y, w, \mathcal{N}, s_j, s_{-j})} dG^{(i)}(\cdot|y, w, \mathcal{N}, \boldsymbol{s}) \tag{49}$$

$$\begin{aligned} B_i(\boldsymbol{s}) &= E_{YW\mathcal{N}} 1\{i \in \mathcal{N}\} \\ &\times \textstyle\sum_{j\in\mathcal{N}, j\neq i} \int \left[1 + \frac{h(\cdot|y,w,\mathcal{N},s_i,s_{-i})}{\sum_{l\in\mathcal{N}, l\neq i} h(\cdot|y,w,\mathcal{N},s_l,s_{-l})}\right] dG^{(j)}(\cdot|y, w, \mathcal{N}, \boldsymbol{s}) \\ &\times (1\{\omega(x, \boldsymbol{s}, j) = s^1\}, \ldots, 1\{\omega(x, \boldsymbol{s}, j) = s^m\}) \end{aligned} \tag{50}$$

$$\begin{aligned} D_i(\boldsymbol{s}) &= E_{YW\mathcal{N}} 1\{i \notin \mathcal{N}\} \\ &\times \textstyle\sum_{j\in\mathcal{N}} \int dG^{(j)}(\cdot|y, w, \mathcal{N}, \boldsymbol{s}) \\ &\times (1\{\omega(x, \boldsymbol{s}, j) = \boldsymbol{s}^1\}, \ldots, 1\{\omega(x, \boldsymbol{s}, j) = \boldsymbol{s}^m\}) \end{aligned} \tag{51}$$

[45] Assumption 5.1 also provides an ordering of the elements of the diagonal of $\mathbf{g_{Eb|W}}$ and columns of $\mathbf{g_{W^B|W}}$ accordingly. Note that the elements of $\mathbf{g_{W^B|W}}$ are probabilities and each column should add to 1, hence I rescale the columns appropriately.

Let $V_i = (V_i(\boldsymbol{s}^1), \ldots, V_i(\boldsymbol{s}^m))'$, $A_i = (A_i(\boldsymbol{s}^1), \ldots, A_i(\boldsymbol{s}^m))'$, $B_i = (B_i(\boldsymbol{s}^1), \ldots, B_i(\boldsymbol{s}^m))'$, and $D_i = (D_i(\boldsymbol{s}^1), \ldots, D_i(\boldsymbol{s}^m))'$. Then I can write equation (11) in matrix form

(52) $V_i = A_i + \beta(B_i + D_i)V_i$

and then solve for V_i to get

(53) $V_i = [I - \beta(B_i + D_i)]^{-1} A_i$

B Estimation of the Value Function

I follow the procedure in JP, and approximate the value function, given in equation (12), on a grid of state vectors. Denote the grid of state vectors $\hat{S} = (\boldsymbol{s}^1, \ldots, \boldsymbol{s}^l)$, and restrict the transition function ω to $\hat{S}$ by defining a transition function $\hat{\omega}(x, \boldsymbol{s}, j) = \{\boldsymbol{s} \in \hat{S} | \boldsymbol{s} \text{ is closest to } \omega(x, \boldsymbol{s}, j)\}$. I select the grid by drawing 200 states from the distribution of observed states.

For every regular bidder, I solve for the value function using (53) on the grid $\hat{S}$. To do so, I numerically evaluate the terms A_i, B_i and D_i given in (49)-(51). In all three expressions, the first expectation is with respect to contract characteristics. Remember that contract characteristics are iid and $Y \perp\!\!\!\perp W$, so we can approximate the expectation with respect to Y using a sample average taking a sample of Y from the observed characteristics (I draw a sample of 500 characteristics). The expectation with respect to W is easy to compute since W is discrete and we have an estimate of its distribution, F_W, from (20). On the other hand, the expectation with respect to $\mathcal{N}$ is more involved. Since I do not estimate a structural model for individual participation decisions, in order to compute the expectation with respect to $\mathcal{N}$ I use the following approximation shortcut. Given Y, W and $\boldsymbol{s}$, I predict the number of bidders in the auction using the estimate of (5) and (19). Using the observed average ratio of the number of regular bidders to the number of total bidders I get a prediction of the number of regular and fringe firms. To select which regular firms participate in the auction, I use a probit model of individual participation and select the required number of regular firms with the highest predicted probability.

Finally, I evaluate the expectation with respect to the bid distribution functions by numerical integration using the estimated derivatives $d\hat{G}^{(i)}(\cdot|y, w, \mathcal{N}, \boldsymbol{s})$ and $d\hat{G}^{(j)}(\cdot|y, w, \mathcal{N}, \boldsymbol{s})$.

To evaluate the value function at points outside the grid $\hat{S}$, I use a quadratic polynomial approximation.

C Density Estimation: Additional Results

$$\hat{\mathbf{g}}_{\mathbf{W^B},\mathbf{W^A}} = \begin{pmatrix} 0.0283 & 0.0349 & 0.0109 & 0.0021 \\ 0.0377 & 0.1794 & 0.0954 & 0.0385 \\ 0.0047 & 0.0792 & 0.1497 & 0.0714 \\ 0.0001 & 0.0059 & 0.0978 & 0.1642 \end{pmatrix}$$

$$\hat{\mathbf{g}}_{\mathbf{W^B}|\mathbf{W}} = \begin{pmatrix} 0.6391 & 0.0798 & 0.0001 & 0.0001 \\ 0.3485 & 0.8681 & 0.0638 & 0.0078 \\ 0.0068 & 0.0521 & 0.8282 & 0.0541 \\ 0.0057 & 0.0001 & 0.1080 & 0.9381 \end{pmatrix}$$

Table 16: Bid Density Estimates

	θ_1^R	θ_2^R	θ_1^F	θ_2^F
constant	-2.1126e+00	-1.4539e+00	-1.5251e+00	-1.7549e+00
log(days)		-4.9912e-03		1.2950e-03
log(eng)	3.2961e-01	2.1089e-01	2.8586e-01	2.2781e-01
log(dist)		5.9775e-05		3.7945e-05
# fringe		-1.2843e-02		-9.5630e-03
# regular		-1.3156e-02		-9.6571e-03
std bl		4.3193e-03		
sum std bl $\in \mathcal{N}$		1.8067e-03		4.1775e-03
sum std bl $\notin \mathcal{N}$		1.1059e-03		1.0170e-03
$W^A = 1$		-1.5315e-02		-1.3095e-02
$W^A = 2$		-6.8367e-03		-3.7954e-03
$W^A = 3$		-2.6212e-03		-5.4175e-03
$W^B = 1$	1.7492e-01	-6.9788e-02	8.2094e-03	-2.0289e-02
$W^B = 2$	1.6193e-01	-5.5952e-02	7.4023e-02	-2.2030e-02
$W^B = 3$	1.8858e-01	-2.6665e-02	1.0775e-01	-9.6095e-03

All specifications include time and district dummies.

D Simulation Details

D.1 The effects of the stimulus package on procurement costs

Instead of modifying the projects themselves, as proposed in Section 9.1 (i.e., splitting them in tiny bit parts), one way to get the results under the counterfactual is by changing the firms' beliefs about the effect of the projects on their backlogs. The strategy I follow in practice is to "kill" the backlog effect on the regular firms' FOC. To do so, I compute the equilibrium strategies for regular firms in the static game. Under this new equilibrium regular firms do not take into account the effect of the current project on the future stream of payoffs. Even though fringe firms are already myopic, I also have to compute their equilibrium strategies as their beliefs regarding the regular firms change. In the asymmetric case, the equilibrium strategies of the static game are the solution to a system of first-order ordinary differential equations with boundary conditions. Bajari (2001) shows the existence and uniqueness of the equilibrium and also provides three algorithms to compute the equilibrium. Furthermore, to be consistent with the firms' beliefs about the effect of stimulus projects on backlogs I do not update regular firms' backlogs when they win a stimulus funded project.

Note that the winning bids in this static game will give me the procurement prices for the stimulus projects as if they did not have any effect on backlogs.

For the ARRA funded projects, and for a fixed value of the unobservable, I use the estimates of the distribution of costs functions obtained from the dynamic model and then solve the FOC of the (asymmetric) static game given by

$$(54) \quad 1 + \sum_{j \neq i} \frac{f_j(\phi_j(b))\phi_j'(b)(b - \phi_i(b))}{(1 - F_j(\phi_j(b)))} \quad = \quad 0, \quad i = 1, \ldots, N$$

where $\phi_i(b)$ is the inverse bid function (in the static game). The subscript i is a shorthand to denote that the inverse bid function depends on the backlogs of all regular firms from the point of view of firm i. Equation (54) shows that the inverse bid functions can be characterized as a system of N ordinary differential equations. To solve for them I use a polynomial approximation to the inverse bid functions (see Bajari (2001) for details). This algorithm has also been used in Bajari and Ye (2000).

Once the bid functions are recovered, I simulate bids for the ARRA projects as follows. For each project I observe Y, an estimate of W^B, and $\boldsymbol{s}$. I also have an estimate of the conditional distribution $\Pr(W|W^B)$, and so I draw W from this distribution. Then I draw a sample of costs from the estimated (conditional) cost distributions and obtain the equilibrium bids for each bidder. Note that these bids will not incorporate the effect of the backlog on future profits as they are coming from the static model. The only effect that is accounted for is the effect that previous backlogs have on the distribution of costs, which is the source of the asymmetry between firms. I determine the price for the project by taking the low bid. I repeat these steps 400 times for each ARRA project and then take the average price.

For projects funded by other sources, I obtain their equilibrium prices and firms' costs using the full dynamic model with the counterfactual series of backlogs by drawing bids from the estimated (conditional) distribution of equilibrium bids.

D.2 Delaying the Stimulus

The exercise I carry out in the simulations is as follows. I compare the low bid (i.e., the equilibrium price paid by the government) and the cost of the winning firm, for all projects (those funded with the stimulus money and those funded from other sources) under three situations. First, the baseline takes the timing of the stream of stimulus projects as observed in the data. Then, I perform two counterfactuals that delay the start of the stream of stimulus projects (while keeping the pace of the projects within the stream as observed in the data) by 3 and 6 months, respectively. The timing of projects funded from other sources is kept unchanged.

Even though I do not observe and cannot recover the true unobserved characteristic associated with any project, I do have an estimate of W^B for each project and an estimate of the conditional probability distribution $\Pr(W|W^B)$. I draw W from the latter conditional distribution and then I simulate bids from the estimated distribution of equilibrium bids. For each project, starting on April 2009 (which is when the first stimulus project was offered), I compute the equilibrium price, recover the cost of the winning firm, and update the backlogs according to the identity of the winner. I repeat this exercise 400 times, and compute average prices and actual costs for each project. It is important to note that changes in the backlogs may generate changes in the number of bidders. I take this into account using the estimated reduced form equation for the number of bidders. For the counterfactual, I do the same exercise but delay the stimulus funded projects by 3 or 6 months.

E Additional Participation Results

Table 17: Number of Regular and Fringe Bidders

	# Regular	# Fringe
constant	-2.118***	18.377***
	(.220)	(.835)
log(eng)	.260***	-1.003***
	(.017)	(.063)
log(days)	-.073**	.288***
	(.026)	(.099)
items	.0005	.017***
	(.0005)	(.001)
potential firms	.0005*	-.0003
	(.0002)	(.0009)
sum std bl	-.030***	-.214***
	(.005)	(.021)
nobs	5164	5164
R^2	0.25	0.19

Dependent variable is the number of bidders. All regressions include time, district, and type of work dummies. Standard errors in parenthesis. ***, **, * denote significance at the 1%, 5% and 10% level.

References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6, chap. 63. Elsevier.

Ackerberg, D., K. Caves, and G. Fraser (2004): "Structural identification of production functions," working paper, UCLA.

An, Y., Y. Hu, and M. Shum (2010): "Estimating first-price auctions with an unknown number of bidders: A misclassification approach," *Journal of Econometrics*, 157, 328–341.

Athey, S., D. Coey, and J. Levin (2011): "Set-Asides and Subsidies in Auctions," Working paper.

Athey, S., and P. Haile (2007): "Nonparametric Approaches to Auctions," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6, chap. 60. Elsevier.

Athey, S., J. Levin, and E. Seira (2011): "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions," *The Quarterly Journal of Economics*, 126, 207–257.

Bajari, P. (1997): "The First-Price Auction with Asymmetric Bidders: Theory and Applications," University of Minnesota Ph.D. thesis.

——— (2001): "Comparing Competition and Collusion: a Numerical Approach," *Economic Theory*, 18, 187–205.

Bajari, P., and A. Hortacsu (2003): "The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights From eBay Auctions," *Rand Journal of Economics*, 34, 329355.

Bajari, P., S. Houghton, and S. Tadelis (2006): "Bidding for Incomplete Contracts: An Empirical Analysis," NBER working paper.

Bajari, P., and S. Tadelis (2001): "Incentives Versus Transaction Costs: A Theory of Procurement Contracts," *Rand Journal of Economics*, 32, 287–307.

Bajari, P., and L. Ye (2000): "Sealed-bid Auctions with Asymmetric Bidders: Computation, Identification and Estimation," Working paper.

——— (2003): "Deciding Between Competition and Collusion," *The Review of Economics and Statistics*, 85, 971–989.

Balat, J., and P. Haile (2011): "Participation Model Estimation: Technical Appendix," Technical note.

CAMPO, S., I. PERRIGNE, AND Q. VUONG (2003): "Asymmetry in First-Price Auctions with Affiliated Private Values," *Journal of Applied Econometrics*, 18, 197–207.

CANTILLON, E. (2008): "The Effect of Bidders' Asymmetries on Expected Revenue in Auctions," *Games and Economic Behaviour*, 62, 1–25.

CHEN, X., H. HONG, AND E. TAMER (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343–366.

CHEN, X., Y. HU, AND A. LEWBEL (2008a): "Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information," *Economics Letters*, 100, 381384.

——— (2008b): "A note on the closed-form identification of regression models with a mismeasured binary regressor," *Statistics and Probability Letters*, 78, 1473–1479.

——— (2009): "Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments," *Statistica Sinica*, 1, 1–3.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

DESILVA, D., T. DUNNE, AND G. KOSMOPOULOU (2003): "An Empirical Analysis of Entrant and Incumbent Bidding in Road Construction Auctions," *The Journal of Industrial Economics*, LI(3), 295–316.

EINAV, L., AND I. ESPONDA (2008): "Endogenous Participation and Local Market Power in Highway Procurement," working paper.

FLAMBARD, V., AND I. PERRIGNE (2006): "Asymmetry in Procurement Auctions: Evidence from Snow Removal Contracts," *The Economic Journal*, 116, 1014–1036.

GOOLSBEE, A. (1998): "Does Government R&D Policy Mainly Benefit Scientists and Engineers?," NBER working paper.

GROEGER, J. (2010): "Participation in Dynamic Auctions," working paper.

GUERRE, E., I. PERRIGNE, AND Q. VUONG (2000): "Estimation of a Dynamic Auction Game," *Econometrica*, 68, 525–574.

——— (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77, 1193–1227.

HAILE, P. A., H. HONG, AND M. SHUM (2006): "Nonparametric Tests for Common Values In First-Price Sealed-Bid Auctions," working paper.

HALL, P., AND X.-H. ZHOU (2003): "Nonparametric identification of component distributions in a multivariate mixture," *Annals of Statistics*, 31, 201–224.

Heckman, J., and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

Henry, M., Y. Kitamura, and B. Salanie (2011): "Identifying Finite Mixtures in Econometric Models," working paper.

Hong, H., and M. Shum (2002): "Increasing Competition and the Winner's Curse: Evidence from Procurement," *The Review of Economic Studies*, 69, 871–898.

Hu, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27–61.

Hu, Y., D. McAdams, and M. Shum (2011): "Identification of First-Price Auctions with Non-separable Unobserved Heterogeneity," working paper.

Hu, Y., and S. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Jofre-Bonet, M., and M. Pesendorfer (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 1443–1489.

Kitamura, Y. (2003): "Nonparametric identifiability of finite mixtures," Unpublished manuscript.

Krasnokutskaya, E. (2011): "Identification and Estimation in Procurement Auctions under Unobserved Auction Heterogeneity," *Review of Economic Studies*, forthcoming.

Krasnokutskaya, E., and K. Seim (2011): "Bid Preference Programs and Participation in Highway Procurement Auctions," *The American Economic Review*, 101, 2653–2686.

Levin, D., and J. Smith (1994): "Equilibrium in Auction with Entry," *American Economic Review*, 84, 585–599.

Levinsohn, J., and A. Petrin (2003): "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies*, 70 (2), 317341.

Lewbel, A. (2007): "Estimation of average treatment effects with misclassification," *Econometrica*, 75, 537551.

Li, T., I. Perrigne, and Q. Vuong (2000): "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 98, 129–161.

——— (2002): "Structural Estimation of the Affiliated Private Values Model," *RAND Journal of Economics*, 33, 171–193.

Li, T., and X. Zheng (2009): "Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions," *Review of Economic Studies*, 76, 1397–1429.

Mahajan, A. (2006): "Identification and estimation of regression models with misclassification," *Econometrica*, 74, 631665.

Marmer, V., A. Shneyerov, and P. Xu (2011): "What Model for Entry in First-Price Auctions? A Nonparametric Approach," Working paper.

Maskin, E., and J. Riley (2000): "Asymmetric Auctions," *Review of Economic Studies*, 67, 413–438.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Porter, R., and D. Zona (1993): "Detection of Bid Rigging in Procurement Auctions," *Journal of Political Economy*, 101, 518–538.

Roberts, J. (2011): "Unobserved Heterogeneity and Reserve Prices in Auctions," working paper.

Roberts, J., and A. Sweeting (2010): "Entry and Selection in Auctions," Working paper.

Saini, V. (2011): "Endogenous Asymmetry in a Dynamic Procurement Auction," Working paper.

Samuelson, W. (1985): "Competitive Bidding with Entry Costs," *Economic Letters*, 17, 53–57.