

Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions*

Tong Li and Xiaoyong Zheng[†]

This Version: January 2007 (First Version: October 2004)

Abstract

Motivated by several interesting features of the highway mowing auction data from Texas Department of Transportation (TDoT), we propose a two-stage procurement auction model with endogenous entry and uncertain number of actual bidders. Our entry and bidding models provide several interesting implications. For the first time, we show that even within an independent private value paradigm, as the number of potential bidders increases, bidders' equilibrium bidding behavior may become less aggressive because the "entry effect" is always positive and may dominate the negative "competition effect." We also show that it is possible that the relationship between the expected winning bid and the number of potential bidders is non-monotone decreasing as well. We then develop a structural model of entry and bidding controlling for unobserved auction heterogeneity and use the recently developed semiparametric Bayesian estimation method to analyze the data. The structural estimates are used to quantify the "entry effect" and the "competition effect" with regard to the individual bids and the procurement cost, as well as the savings for the government with regard to the procurement cost when the entry cost is reduced.

*We thank S. Athey, M. Baye, R. Becker, R. Blundell, A. Chesher, W. Collins, Y. Fan, R. Gardner, P. Haile, K. Hendricks, A. Hortaçsu, E. Krasnokutskaya, D. Levin, H. Paarsch, A. Pakes, M. Pesendorfer, H. Pesaran, P. Schimdt-Dengler, A. Shneyerov, and participants in seminars at Harvard, Indiana, LSE, NC State, Rice, University of Chicago, University of Miami, University of Michigan, University of Texas at Austin, UCF, UCL, UCSB, UBC, USC, Vanderbilt and at the Econometric Society World Congress in London, August 2005, the CIREQ Conference on Auctions in Montreal, October 2005, and the Econometric Society Winter Meeting in Boston, January 2006 for helpful discussions and comments. We also thank Texas Department of Transportation, especially, R. Smith for making the data available. The first author would like to thank CEMMAP at UCL and its director Professor Andrew Chesher for their hospitality during the completion of this paper. Previous versions of the paper were circulated under the title "Procurement Auctions with Entry and Uncertain Number of Actual Bidders: Theory, Structural Inference and an Application."

[†]Li: Department of Economics, Vanderbilt University, tong.li@vanderbilt.edu; Zheng: Department of Agricultural and Resource Economics, North Carolina State University, xzheng@ncsu.edu.

1 Introduction

In this paper, motivated by several interesting features of the highway mowing auction data we collected from Texas Department of Transportation (TDoT), we propose a two-stage procurement auction model with endogenous entry and uncertain number of actual bidders. We then develop a fully structural econometric framework that is derived from our game-theoretic model of entry and bidding, and that also allows for controlling for unobserved auction heterogeneity. We develop a semiparametric Bayesian method to estimate the underlying structural elements, namely, the entry cost distribution, the bidders' private cost distribution, and the distribution of the unobserved heterogeneity, and apply it to conduct a detailed structural analysis of the procurement auction data from TDoT.

Our theoretical model of entry and bidding is motivated by the strong evidence of entry behavior in the data set: on average only about 28.05% of the potential bidders (those who request for official bidding proposals) actually submit their bids. Endogenous entry because of the participation cost such as the cost involved in acquiring information and preparing for bidding as well as the opportunity cost in bidding (and winning) in one auction has been documented in recent empirical work (Bajari and Hortacısu (2003) for eBay auctions, Athey, Levin and Seira (2004) for timber auctions). Theoretical models of entry, on the other hand, have been developed since 1980s with Samuelson (1985) and McAfee and McMillan (1987a) studying pure strategies of entry, and Levin and Smith (1994), Harstad (1990), Kjerstad and Vagstad (2000), as well as McAfee, Quan and Vincent (2002) studying mixed strategies. As in Levin and Smith (1994), we assume that each potential bidder adopts a symmetric mixed strategy in deciding whether to incur a cost in order to acquire his signal and submit a bid.¹ Our model, on the other hand, differs from Levin and Smith (1994) in that while Levin and Smith (1994) assume that actual bidders know the number of actual bidders at the time of bidding, we relax this assumption.² We derive some interesting model implications, most of which are empirically testable. For the first time, we show that even within the independent private value (IPV) paradigm, as the number of potential bidders increases, bidders' equilibrium bidding behavior can become less aggressive. Thus, increasing competition may not be always desirable for the government, meaning that our result, while somewhat counter-

¹A pure strategy entry model such as Samuelson (1985), in contrast, assumes that potential bidders first draw their signals and then decide whether to incur a cost to prepare a bid.

²In contrast to the previous work that assumes (exogenously) varying number of actual bidders and studies revenue and welfare implications (McAfee and McMillan (1987b), Matthew (1987), Harstad, Kagel and Levin (1990), Levin and Ozdenoren (2004)), the uncertain number of actual bidders in our model is endogenous as a result of the endogenous entry.

intuitive and surprising because of the pure IPV paradigm under consideration, can have important policy implications. This result can also be used to test empirically whether entry is an important part of the decision making process. Notably, the relationship between bids and number of potential bidders has been an important issue pursued in the theoretical and empirical literature, but has been limited to the framework in which there is no entry. In this framework, it has been shown that while in the IPV paradigm the relationship is monotone, it may not be in a common value (CV) model because of the interaction between the “competition effect” and “winner’s curse effect” (Bulow and Klemperer (2002)), and in an affiliated private value model due to the opposite “competition effect” and “affiliation effect” (Pinkse and Tan (2004)). Our result is driven by the interaction of two opposite effects: the “competition effect” and the “entry effect.” While the “competition effect” is always negative as usual, the “entry effect” is always positive. This positive “entry effect” suggests to a winning bidder that he may have overestimated the intensity of entry.

Another interesting question is whether the “entry effect” and the “competition effect” remain the same sign respectively for the expected winning bid (i.e. procurement cost) as in the individual bid case. To this end, we provide a sufficient condition for the “entry effect” to be positive and for the “competition effect” to be negative for the expected procurement cost. The condition is general and met by most of the distributions that are commonly used in modeling private cost. As a result, we show that it is possible for the procurement cost to be non-monotone decreasing in the number of potential bidders, as this can be the case when the “entry effect” is positive and the “competition effect” is negative and the former dominates the latter.³ We also show that under the same sufficient condition the procurement cost decreases with the entry cost.

To analyze the data we have collected and also to provide a general framework within which

³In studying high-bid auctions with entry assuming that actual bidders know the number of actual bidders at the time of bidding, Levin and Smith (1994) show that under the optimal mechanism for the auctioneer where the auctioneer’s expected revenue is maximized, and hence the optimal entry is induced, the expected revenue (same as the expected winning bid) becomes less aggressive with increasing number of potential bidders when the number of potential bidders grows beyond some cut-off point. Our result, on the other hand, does not require the optimal mechanism for the auctioneer, and provides sufficient conditions under which the expected winning bid may become less aggressive with the number of potential bidders. Thus our result on the relationship between the procurement cost and the number of potential bidders complements Levin and Smith’s result, while our result on the relationship between the individual bids and the number of potential bidders is the first established in the literature, to the best of our knowledge.

entry and bidding are jointly modeled, we develop a fully structural model jointly modeling entry and bidding based on the theoretical model we propose. Our structural approach takes into account of controlling for unobserved auction heterogeneity, which is necessary as in procurement auctions, bidders have access to the information about the auctioned projects, some of which may not show up in the data.⁴ From an econometric viewpoint, failing to control for the unobserved auction heterogeneity in empirical research could severely bias the estimates for the structural parameters, and hence yield misleading policy conclusions and recommendations, as shown in Krasnokutskaya (2002).

It is worth noting that joint modeling of entry and bidding is on one hand a modeling issue and on the other hand challenging from the perspective of econometric implementation, while controlling for unobserved auction heterogeneity is mainly an important econometric issue. This paper aims to provide a unified framework to address these two issues in structural inference of auctions. We achieve this objective by parameterizing the bidders' private cost distribution and the entry cost distribution, leaving the distribution of the unobserved heterogeneity term unspecified, and adopting the recently developed semiparametric Bayesian method to estimate the underlying structural elements. The semiparametric Bayesian method using a Dirichlet process prior is introduced by Lo (1984) and Ferguson (1983), with its recent development stimulated by the advances in the Markov chain Monte Carlo (MCMC) method (e.g. Escobar (1994), Escobar and West (1995)). While this method has been recently adopted by econometricians in addressing various issues (Hirano (2002) for estimation of linear autoregressive panel data models, Chib and Hamilton (2002) for analysis of longitudinal data treatment effect models, Hasegawa and Kozumi (2003) for estimation of Lorenz curves, and Griffin and Steel (2004) for inference of stochastic frontier models), our paper demonstrates its considerable advantage in making complicated structural models tractable.⁵

Jointly estimating fully structural models of entry and bidding has been a challenging problem because of the complexity involved with the joint models of entry and bidding, as well as the presence of latent variables. We are able to circumvent the difficulties by using the data augmentation

⁴For example, in our data set we do not observe the information on road conditions that may affect bidders' decision.

⁵Bajari (1997) first uses parametric Bayesian methods to estimate asymmetric auction models. Subsequent work by Bajari and his coauthors also uses parametric Bayesian methods to estimate auction models. Also, Sareen (1999) compares the IPV and the CV paradigms using posterior odds comparison.

techniques that enable us to develop a full Monte Carlo Markov chain to estimate the parameters in the structural model and the distribution of the unobserved heterogeneity. This contributes to the literature of structural analysis of auction data, as to the best of our knowledge, our paper is the first one that estimates a fully structural model for entry and bidding, while controlling for unobserved auction heterogeneity.⁶

We use our structural approach to analyze the TDoT data. Our results indicate that there is a significant effect from the unobserved heterogeneity. We find that the entry cost is a significant part in the bidders' decision making process and hence a theoretical model or empirical analysis that ignores the entry effect may lead to poor policy recommendations.⁷

After obtaining the structural elements of the model, we then conduct counterfactual analysis by quantifying the effects of number of potential bidders on the individual bid as well as on the procurement cost. We find that the "entry effect" is positive and the "competition effect" is negative. Moreover, the positive "entry effect" actually dominates the negative "competition effect" in our data and hence policies that encourage more potential bidders might cause the government's procurement costs to increase. We also quantify the savings for the government with regard to the procurement cost when the entry cost is reduced.

The empirical analysis of the procurement auctions conducted in this paper contributes to the growing literature on empirical analyses of procurement auctions using structural approach. For example, see Bajari and Ye (2002) for detecting collusion, Hong and Shum (2002) for assessing the winner's curse, Jofre-Bonet and Pesendorfer (2003) for capacity constraint in dynamic procurement

⁶Bajari and Hortacısu (2003) as well as Athey, Levin and Seira (2004) are important developments in taking into account bidders' participation when using the structural approach to analyze auction data, both using a reduced form specification for modeling entry while estimating a structural bidding model. Li (2005) derives moment conditions implied by the entry and bidding model in Levin and Smith (1994) when reserve prices are binding, which can be used for estimation and testing. He also suggests to use a flexible reduced form specification for the entry probability. On the other hand, Haile, Hong and Shum (2002) develop tests for common values in first-price auctions allowing for entry and unobserved auction heterogeneity.

⁷As in Athey, Levin and Seira (2004) and Bajari and Hortacısu (2003), we are able to recover the entry cost in dollar terms using the equilibrium conditions and structural estimates. As a result, although in a different information environment with a different game structure, the structural approach in entry and bidding of auctions allows one to recover firms' profit functions in a similar spirit to the pioneering work by Bresnahan and Reiss (1990) and Berry (1992) who study entry in monopoly/oligopoly markets.

auctions, Krasnokutskaya (2002) for the effect of unobserved auction heterogeneity in asymmetric first-price IPV auctions, and Krasnokutskaya and Seim (2006) for bid preference programs and participation in highway procurement auctions.⁸ Most of the work, while studying different aspects of procurement auctions, however, has assumed exogenous participation; a notable exception is Krasnokutskaya and Seim (2006). Our empirical analysis, on the other hand, is based on a fully structural model for entry and bidding, and quantifies the effect of entry in the bidders' decision process, as well as the role of the unobserved auction heterogeneity. In particular, the structural approach adopted here enables us to assess the "entry effect" and the "competition effect," as well as to quantify the relationship between the procurement cost and the entry cost, and hence can be useful for making policy recommendations.

This paper is organized as follows. Section 2 presents the data set that will be analyzed to motivate the problem, and to discuss the relevance of the theoretical model later developed. Section 3 is devoted to proposing a mixed strategy entry model along with a bidding model, and deriving some model implications from the joint entry and bidding model. Section 4 carries out several reduced-form empirical tests of the theoretical model using our data. Section 5 develops the structural framework for joint modeling of entry and bidding from the theoretical model proposed in Section 3 and proposes the semiparametric Bayesian Markov chain Monte Carlo (MCMC) estimation algorithm to estimate the structural model. Estimation results are presented in Section 6. Section 7 carries out counterfactual analyses to quantify the "entry effect" and the "competition effect" on the individual bid and the procurement cost as well as the savings for the government when the entry cost is reduced. Section 8 concludes. The technical proofs and algorithms are included in Appendix.

2 Data

The data used in this study is from the highway construction and maintenance procurement auctions held by the Texas Department of Transportation (TDoT) between January 2001 and December 2003. We focus on a particular type of highway maintenance job, which is called "mowing highway right-of-way." We select this type of auctions for two reasons. First, this type of job is the single most

⁸Another strand is to use reduced-form tests to test collusion in procurement auctions. Porter and Zona (1993) represents this line of research.

frequently held auction in the sample period we study. Second, this type of auctions is relatively simple compared with other big construction projects. These highway mowing projects usually involve a main task, which is called “type-II full width mowing,” with several additional tasks such as strip mowing, spot mowing, litter pickup and disposal, sign installation and so on.

The TDoT advertises projects 3 to 6 weeks prior to the letting date. Advertisement usually includes a short description of the project including the location, the completion time, the engineer’s estimate and the estimated quantity for each task involved. In order to become an eligible bidder for a certain project, the contractor must first become a qualified bidder, which is based on the contractor’s financial statement and is reviewed on a yearly basis.⁹ TDoT maintains a master list for these qualified bidders, which lists all the qualified bidders in any sub-industry of the highway construction and maintenance industry in Texas. Second, those qualified contractors interested in the project must obtain a detailed project plan and the official bidding proposal from the TDoT no later than 21 days prior to the letting date. TDoT maintains a bidders’ list of those contractors who have requested the official bidding proposal for the project. We observe from the data that for a certain project, only (and usually all of) those contractors in the mowing sub-industry who are located in the same county as where the job is or nearby counties would request the official bidding proposals. Therefore, we treat the bidders on the bidders’ list for a certain project rather than the bidders on the master list as the potential bidders. Bids have to be submitted before the bid opening time in a sealed envelope.

The information flow of the auction mechanism is as follows. First, after a project advertisement is posted, bidders learn the auction specifics. During 21 days prior to the bid submission deadline, bidders learn how many potential bidders have already requested the official bidding proposal. Then they prepare and submit bids by the bid opening time without the knowledge of the number of actual bidders. Finally, all the bids and actual bidders’ identities are revealed after the bid opening.

The data set consists of 553 projects, with a total of 1606 bids. The information we observe from the data includes the letting date, the location (district, county, and highway), the tasks involved, the quantity of each task, the identity of all the planholders (potential bidders on the list of bidders), the identity of all the bidders who actually submit their bids (actual bidders), the actual bidding

⁹For certain types of small projects like the mowing job studied in this paper, a contractor does not need to submit the financial statement to become qualified. Instead, they submit a bidder questionnaire to become qualified.

amount, the completion time, the amount of guaranty money, and whether it is a state or local maintenance job.

Table 1 gives the summary statistics of the data and several other quantities of interest. First, it is worth noting that the entry behavior is present in our data set. For the auctions in our data set, on average, an auction has about 11 potential bidders, but the number of actual bidders is only about 2.90. That only 28.05% of the potential bidders actually submit their bids later is a strong evidence of bidders' entry behavior. While requesting a bidding proposal is free, acquiring signals and preparing for a bid is not. The entry behavior indicates that bidders rationally take the entry cost into account and decides whether to enter into the final bidding process.

The mean of the engineer's estimate for these auctions is \$165,348.90, which is very close to the mean of the bids, that is, \$165,382.20. This shows that the engineer's estimate is the key determinant of bidders' bids. On average, these projects last for about four months. Naturally, longer duration of the project will increase bidders' bids for the project. The size of the main task, that is, the "type-II full width mowing," is 7,302.94 acreage on average, with an average of 1,987.64 acreage of other mowing jobs. If bidders have different costs for completing the two different kinds of jobs, then different composition of the jobs will lead bidders to submit different bids even though the total acreage of the mowing jobs are similar. The mean number of tasks in one mowing job is 2.01. Around 11% of the jobs are auctioned by the state agency and 23% of the jobs are on an interstate highway. We distinguish the state job from the local job because we expect that the two agencies may have different requirements for bids preparation and hence the entry cost may differ and hence the bidders change their bids accordingly. Also, a job on an interstate highway may cause bidders to bid more since it is both hard to transport equipment to an interstate highway and to complete the job on an interstate highway due to the high volume of the traffic.

Although the TDoT states in the bidding rules that it will reject the contract if "the lowest bid is higher than the Department's estimate and re-advertising the project for bids may result in lower bids," in reality (at least in the auctions in our data), we do not see that this rule is enforced. In our data set, we observe 704 bids higher than the Department's estimate, which account for about 43.86% of the entire sample. For 120 projects the winning bid is higher than the Department's estimate. These facts suggest that this rule has been ignored by bidders when making their decisions as it has never been implemented. The assumption of no binding public reserve price

is justified in this environment. In effect, the auctions in our study can be considered as first (low)-price sealed-bid auctions without binding public reserve prices with the project awarded to the bidder with the lowest bid.

Lastly, there are about 13 auctions in our data having only one actual bidder. This accounts for about 2.35% of the entire sample. Our entry and bidding model will allow for the presence of only one actual bidder, as can be seen later.

3 Game-Theoretic Framework for Entry and Bidding

In this section, we propose a new game-theoretic model of entry and bidding in procurement auctions and derive its equilibrium properties.

3.1 The Underlying Structure

The government lets a single and indivisible contract to N firm contractors who request the official bidding proposals and become potential bidders. While our model and econometric method can be readily adopted to the case where government announces a reserve price before the auction and the reserve price is effectively binding, we will focus on auctions without reserve prices to be consistent with the data under study. Each potential bidder is risk-neutral with a disutility equal to his private cost c . Each potential bidder must incur an entry cost k to learn his private cost for completing the job and hence how to bid for the project. Thus, the auction is modeled as a two-stage game.

At the first stage, knowing the number of potential bidders N , each potential bidder learns the specifications of the project and the entry cost, calculates their expected profit from entering the auction conditioning on his winning, and then decides whether or not to participate in the auction and actually submit a bid. Moreover, all potential bidders do not draw their private costs until after they decide to enter the auction.

After the first entry stage, the n firm contractors who decide to enter the auction learn their own costs of completing the job. The cost of completing the contract c to a contractor is drawn from a distribution $F(\cdot)$ with support $[\underline{c}, \bar{c}]$. $F(\cdot)$ is twice continuously differentiable and has a density $f(\cdot)$ that is strictly positive on the support. In the independent private value paradigm, when forming his bid, each actual bidder knows his private cost c , but does not know others' private costs. On

the other hand, each bidder knows that all private costs are independently drawn from $F(\cdot)$, which is a common knowledge to all bidders. As a result, all bidders are identical *a priori* and the game is symmetric.¹⁰

Finally, since we assume that there is no binding reserve price, as consistent with the TDoT data, and at the same time, we do observe some auctions with only one actual bidder, it is important to take these two aspects into account to rationalize bidders' behaviors. In first (high)-price sealed-bid auctions without (binding) public reserve prices, there is a unique finite Bayesian-Nash equilibrium bidding strategy with uncertain number of actual bidders. It is not the case, however, for the first (low)-price sealed-bid auctions as considered here. This is because for any rational bidder, if he knows that there is no binding public reserve price, and there is a non-zero probability for him to be the single actual bidder, no matter how small this probability is, his optimal strategy is always to bid infinity. Such a striking difference between high-bid and low-bid auctions, while important, has not been noticed in the literature. Therefore, to rationalize bidders' behaviors so that we can preserve a finite strictly increasing Bayesian-Nash equilibrium bidding strategy in our case, we assume that each potential bidder has a common belief that if he turns out to be the only actual bidder in the auction, he has to compete with the government who will draw its private cost from the same distribution $F(\cdot)$ and form its bid. Otherwise, if he turns out to be one of at least two actual bidders, he just competes with the other actual bidder(s). As a result, it is a common belief for each potential bidder that at each auction, if he decides to enter, he will face at least one another actual bidder, either the government if he is the only actual real bidder or another real bidder if there are at least two actual real bidders.¹¹

¹⁰The assumption that bidders are symmetric can be justified as follows. First, in our data, the winning rate is roughly the same across all bidders. Second, bidders for the mowing jobs are usually individual contractors or 2 or 3-person small firms and hence the difference between bidders are small in terms of efficiency, capacity and other production factors.

¹¹This assumption can be justified because the government does know the number of actual bidders, and also because we do observe auctions with only one actual bidder whose bid is finite. Without introducing a role played by the government, in the absence of a binding public reserve price, these finite bids from single actual bidders cannot be rationalized. In effect, our assumption that the government competes with the single actual bidder is equivalent to introducing a random reserve price from the government when there is a single actual bidder. On the other hand, our assumption is not restrictive as it limits the government's role only to the auctions with a single actual bidder. Furthermore, this assumption is not essential for the theoretical results, as the same results can be shown to hold

3.2 First-Stage: Mixed Strategy Entry

Denote $E\pi(b, c|q^*)$ as the payoff for the actual bidder who optimally bids b using a Bayesian-Nash equilibrium strategy given his own cost c , the unique equilibrium entry probability q^* and the belief that there must be at least two bidders (including himself) in the auction, as well as conditioning on that he is the winner. The ex ante expected payoff for a potential bidder without knowing c is

$$\int_{\underline{c}}^{\bar{c}} E\pi(b, c|q^*)f(c)dc \quad (1)$$

From a game-theoretic viewpoint, a rational bidder will participate in the auction only when the ex ante expected payoff exceeds the corresponding entry cost. Otherwise, remaining outside of the auction is an optimal strategy. In the game-theoretic literature, a mixed strategy model is used based on the assumption that each player adopts a probability distribution for his actions. In our context, the action here refers to the potential bidders' decision on whether or not to enter the auction. Extending Levin and Smith (1994)'s symmetric mixed strategy entry model for high-bid auctions, we assume that there is a unique probability q^* such that all the potential bidders have the same probability q^* to enter the auction. q^* is determined from the equilibrium where the ex ante expected payoff is equal to the entry cost k as given in the following equation¹²

$$\int_{\underline{c}}^{\bar{c}} E\pi(b, c|q^*)f(c)dc = k. \quad (2)$$

3.3 Second-Stage: Bidding

Following the entry stage using the mixed strategy described above, each actual bidder learns his private cost c for completing the job. Different from Levin and Smith (1994), we assume that each

for the low-bid auction case with a binding public reserve price, or the high-bid auction case with/without a binding public reserve price.

¹²There may be other asymmetric pure-strategy equilibria of this two-stage entry and bidding model. We focus on the symmetric mixed-strategy equilibrium for two reasons. First, in our data, in the same region, the participation rate across all the potential bidders is roughly the same and the set of actual bidders changes from auction to auction, indicating that bidders might use the same entry probability to make the entry decision. Second, in our data, for a given set of projects with the same number of potential bidders and roughly the same engineer's estimate, there is still big variation in the number of actual bidders. For example, there are 12 auctions with 11 number of potential bidders and with an engineer's estimate between \$90,000 and \$110,000. The number of actual bidders varies from 2 to 6. This is evidence consistent with the random realization of the binomial process, which is consistent with the mixed-strategy equilibrium.

actual bidder does not know the number of actual bidders at the time of bidding. Taking into account this uncertainty is more relevant to the real applications. Moreover, as in the entry stage, an actual bidder (say bidder i) calculates his bid with the belief that there will be at least two actual bidders in the auction. As a result, his expected profit by bidding b_i conditioning on his private cost c_i and winning the auction is

$$\begin{aligned}
E\pi(b_i, c_i|q^*) &= \sum_{j=2}^N P_B(n = j|n \geq 2)(b_i - c_i) \Pr(b_t \geq b_i, \forall t \neq i) \\
&= (b_i - c_i) \sum_{j=2}^N P_B(n = j|n \geq 2) \Pr(c_t \geq s^{-1}(b_i|q^*), \forall t \neq i) \\
&= (b_i - c_i) \sum_{j=2}^N P_B(n = j|n \geq 2)[1 - F(s^{-1}(b_i|q^*))]^{j-1}, \tag{3}
\end{aligned}$$

where $s(\cdot|q^*)$ is the strictly increasing equilibrium bidding strategy given the entry probability q^* , and

$$P_B(n = j|n \geq 2) = \frac{\binom{N-1}{j-1}(q^*)^{j-1}(1 - q^*)^{N-j}}{1 - (1 - q^*)^{N-1}} \tag{4}$$

is the probability that there will be j actual bidders in the auction from the actual bidder's point of view. Here we make the assumption that bidders have the belief that there will be at least two bidders (including himself) in the auction as discussed in Section 3.1, which is different from Levin and Smith (1994).

The Bayesian-Nash equilibrium of this model can be characterized by the following first-order condition

$$\begin{aligned}
&\frac{\partial \{s(c|q^*) \sum_{j=2}^N P_B(n = j|n \geq 2)[1 - F(c)]^{j-1}\}}{\partial c} \\
&= -c \sum_{j=2}^N P_B(n = j|n \geq 2)(j-1)[1 - F(c)]^{j-2} f(c). \tag{5}
\end{aligned}$$

The unique solution to (5) subject to the boundary condition $s(\bar{c}) = \bar{c}$ is given as follows

$$s(c|q^*) = c + \frac{\sum_{j=2}^N [P_B(n = j|n \geq 2) \int_c^{\bar{c}} (1 - F(x))^{j-1} dx]}{\sum_{j=2}^N [P_B(n = j|n \geq 2)(1 - F(c))^{j-1}]}. \tag{6}$$

As a result, the equation determining the mixed strategy probability q^* (2) can be rewritten as

$$\int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N [P_B(n=j|n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx] f(c) dc = k. \quad (7)$$

3.4 Model Implications

Our entry model for low-bid procurement auctions is similar in spirit to the one for high-bid auctions in Levin and Smith (1994) in that both use mixed strategies to model entry with the difference that ours assumes that bidders make the decision with the belief of having at least two entrants while Levin and Smith (1994) does not. On the other hand, our bidding model is different from Levin and Smith (1994) because we relax the assumption made in Levin and Smith (1994) that the actual bidders know the number of actual bidders at the time of bidding. As a result, our resulting entry model departs significantly from Levin and Smith (1994) because the entry equation (2) is closely associated with the expected payoff of the bidder, which in turn is related to the Bayesian-Nash equilibrium bidding strategy given by (6) from the bidding model. Given such a significant difference, it would be interesting to study the implications from our equilibrium model for entry and bidding, and use them to test whether the entry and bidding behaviors found in our data are consistent with our model. To this end, we will study several comparative statics, most of which are empirically testable.

The first interesting relationship is the one between the entry probability q^* and the entry cost k .

Proposition 1 *In our model, $\frac{\partial q^*}{\partial k} < 0$. That is, as the entry cost increases, holding everything else fixed, the bidders' equilibrium entry probability decreases.*

The negative relationship between k and q^* is intuitive since as the entry cost increases, bidders require higher expected payoff to be compensated for entering into the auction. Therefore, the probability that a bidder can gain from the auction decreases, resulting in less probability of their entering into the auction. Levin and Smith (1994) demonstrate the same relationship in their model.

Before we turn to the relationship between the equilibrium bids b and the number of potential bidders N , we first take a look at the relationship between the entry probability q^* and the number of potential bidders N . This is the key intermediate step in how the number of potential bidders affects bidders' equilibrium bids. From equation (6) for the equilibrium bidding strategy, it is

obvious that when the number of potential bidders increases, it affects the equilibrium bids through two channels. First, it affects the bidding directly. The second effect is through the equilibrium entry probability q^* . When the number of potential bidders changes, q^* changes because they both appear on the left-hand side of equation (2) for entry while the entry cost on the right hand side of the equation is fixed. Despite the complexity of our entry model, we are able to show that the relationship between q^* and N is negative while keeping everything else constant without imposing any conditions.¹³ Note that proving the relationships $\partial q^*/\partial N < 0$ and later $\partial b/\partial N < 0$ is non-standard and a bit technically involved because of the appearance of N on the upper bound of the sum of probabilities, which prevents us from directly taking derivatives.

Proposition 2 *In our model, $\frac{\partial q^*}{\partial N} < 0$. That is, as the number of potential bidders increases, holding everything else fixed, the bidders' equilibrium entry probability decreases.*

The intuition behind this relationship is as follows. q^* is determined by the entry equation (7). On the right hand side, the entry cost k does not change. Therefore, the left-hand side, which is the expected gain for a bidder who enters this auction, also remains the same. In other words, the number of actual bidders the auction can support does not change. As a result, when the number of potential bidders increases, bidders lower their equilibrium entry probability.

Now we turn to the relationship of one of our main interests, that is, the relationship between the equilibrium bid b and the number of potential bidders N . This is an important issue, as previously in the literature with the assumption that the number of actual bidders is known to the actual bidders when they submit bids, the equilibrium bid in low-bid auctions within the IPV paradigm is shown to be monotone decreasing in the number of actual bidders, or in the number of potential bidders. This result has been considered as an important property of IPV model that is often used in empirical research to validate the model. Our model, however, is different because we assume that the actual bidders do not know the number of actual bidders when submitting bids. We investigate the relationship between bid and N analytically. This relationship, if established, can shed light on studying other auction data. In particular, the predicted relationship between bid and N from our

¹³Levin and Smith (1994) give a sufficient condition under which the relationship between the induced entry probability from the optimal auction mechanism and N is negative while everything else is kept constant. Our result gives such a relationship for any entry probability (not necessarily induced by the optimal auction mechanism) without imposing sufficient conditions.

model, which is different from the one with bidders knowing the number of actual bidders, can be used to form a basis for testing whether bidders know the number of actual bidders when submitting bids.

Proposition 3 *In our model, the relationship between b and N may not be monotone decreasing.*

Roughly speaking, this proposition is obtained by rewriting the equilibrium bidding strategy equation (6) as $b = p(N, q^*)$, and taking the total differentiation of b with respect to N

$$\frac{db}{dN} = \frac{\partial b}{\partial N} + \frac{\partial b}{\partial q^*} * \frac{\partial q^*}{\partial N}.$$

This equation illustrates that the number of potential bidders, N , affects the equilibrium bids in two channels. The first channel is the usual “competition effect,” as represented by $\frac{\partial b}{\partial N}$, which is negative as proved by Lemma 3 in the appendix. This is because as the number of potential bidders increases, if the bidders’ equilibrium entry probability is fixed, bidders become more aggressive. The intuition behind this is that when there are more potential bidders but the bidders still use the same equilibrium entry probability to decide whether or not to enter into the auction, the expected number of actual bidders increases. Therefore, actual bidders face more competition and they bid more aggressively. The second channel, which we call the “entry effect,” directs the effects of N on b through the entry stage. As the number of potential bidders increases, holding everything else constant, the bidders’ equilibrium entry probability decreases, as shown in Proposition 2. On the other hand, as the entry probability decreases, holding everything else constant, the equilibrium bidding strategy increases. This is proved in Lemma 4 in Appendix. This result is intuitive as when the number of potential bidders is fixed but the bidders’ equilibrium entry probability decreases, the expected number of actual bidders decreases. Therefore, actual bidders face less competition and they bid less aggressively. The resulting entry effect is thus positive. In the end, the relationship between b and N depends on the relative magnitudes of the “entry effect” and the “competition effect.”

We now illustrate the non-monotone decreasing relationship between b and N by a concrete example. We choose the private cost distribution to be the *uniform*(0,1) and $k = 0.2$. The equilibrium bidding strategies for cases $N = 8$ and $N = 3$ are computed and plotted in Figure (1). Note that when the private production cost c is greater than 0.09, the bidding line for $N = 8$ lies

above the one for $N = 3$. Hence the equilibrium bids are not monotone decreasing in N . Also, consistent with Proposition 2, when $N = 8$, the equilibrium entry probability is 0.26 and it increases to 0.89 when $N = 3$.

An interesting question is whether Proposition 3 holds also for the relationship between the expected winning bid (procurement cost) and the number of potential bidders. While we are able to show that for an individual bid, the “entry effect” is always positive and the “competition effect” is always negative, assessing the signs of the “entry effect” and the “competition effect” is much more involved in the case of the procurement cost. Nevertheless, we are able to provide sufficient conditions for the “entry effect” to be positive and for the “competition effect” to be negative with regard to the procurement cost, which are given in the following proposition.

Proposition 4 *Denote the expected winning bid (procurement cost) by W . If $c + F(c)/f(c)$ is increasing in c , then*

- (i) *the competition effect $\frac{\partial W}{\partial N} < 0$;*
- (ii) *$\frac{\partial W}{\partial q^*} < 0$. As a result, the entry effect $\frac{\partial W}{\partial q^*} * \frac{\partial q^*}{\partial N} > 0$.*
- (iii) *the relationship between W and N may not be monotone decreasing.*

Note that the sufficient condition given here for the competition effect to be negative and for the entry effect to be positive is general as it is satisfied by uniform, exponential, lognormal, logistic, and Weibull (with a shape parameter ≤ 1).¹⁴ Thus for most of the distributions of non-negative valued random variables that are widely used in empirical analysis of auctions, the entry effect is positive and the competition effect is negative. Thus, it is possible that the relationship between the procurement cost and the number of potential bidders is non-monotone decreasing.

As expressed in Lemma 5 in Appendix, the expected winning bid W is a function of q^* , N , and $F(\cdot)$. A direct consequence of Proposition 1 and (ii) in Proposition 4 is that under the same condition as in Proposition 4, as the entry cost decreases, the procurement cost decreases as well.

¹⁴This condition is analogous to the condition used in the optimal auction mechanism literature that $v - (1 - F(v))/f(v)$ is increasing in v in high-bid auctions where v is a bidder’s private value (signal), which defines the design problem to be *regular*. See, e.g. Krishna (2002). The quantity $v - (1 - F(v))/f(v)$ is referred to as the bidder’s virtual valuation or the bidder’s marginal revenue (see, e.g. Bulow and Klemperer (1996) and Krishna (2002)). Similarly, the quantity $c + F(c)/f(c)$ in our case can be considered as the bidder’s virtual cost, and our sufficient condition gives rise to the *regular* design problem in the optimal procurement auction case.

This result is given in the following proposition.

Proposition 5 *When $c + F(c)/f(c)$ is increasing in c , $\frac{\partial W}{\partial k} > 0$.*

Propositions 4 and 5 have important policy implications. They establish the relationships between the expected procurement cost and the number of potential bidders as well as the entry cost, respectively. Since at the equilibrium, the social welfare is equal to the social value of the project minus the expected procurement cost because the expected payoff for a bidder is zero, Proposition 4 implies that it is possible that the social welfare can be worse off as the potential competition becomes more intense, while Proposition 5 indicates that in general the social welfare improves with the entry cost reduced.

4 Reduced-Form Empirical Tests of the Model

4.1 The Relationship between q^* and N

First, we use our procurement auction data to test whether there is a negative relationship between the equilibrium entry probability q^* and the number of potential bidders N , as predicted in Proposition 2.

While q^* is a conceptually defined random variable which is unobserved in our data, we can use the information on the number of actual bidders and the number of potential bidders to infer the relationship between q^* and N . Since n , the number of actual bidders, is a count variable, we assume that the number of actual bidders satisfies the following conditional mean condition

$$E(n|\mathbf{x}, N) = \exp(\mathbf{x}\zeta + \rho \log N) = N^\rho \exp(\mathbf{x}\zeta) \quad (8)$$

as in a Poisson regression model or generalized count regression models. Under this conditional mean assumption, using the Poisson quasi-MLE approach one can consistently estimate ρ as well as ζ . Since (8) implies

$$E\left(\frac{n}{N}|\mathbf{x}, N\right) = N^{\rho-1} \exp(\mathbf{x}\zeta), \quad (9)$$

if the estimated ρ is less than 1, then as N increases, $\frac{n}{N}$, which is a good proxy for q^* , tends to decrease.

We include the logarithm of the engineer’s estimate, logarithm of the number of working days, logarithm of the acreage of the type-II full width mowing, logarithm of the acreage of other types of mowing jobs, the number of items in the job, whether the job is a state job and whether it is a highway job as the set of explanatory variables \mathbf{x} in the Poisson regression model. Estimation results for the parameters ζ and ρ are reported in Table 2.

First, we note that the coefficient on the number of potential bidders is significant with a value of 0.4011, which is strictly less than 1. This result can be viewed as a support of the inverse relationship between q^* and N . As the number of potential bidders increases, the equilibrium entry probability decreases.

Second, since the theoretical model also predicts that the equilibrium entry probability is decreasing in the entry cost k as given in Proposition 1, those explanatory variables that are likely to affect the entry cost may also be significant in the regression. For example, the number of items in the job has a significant negative effect on the number of actual bidders. This may be due to the fact that more items a job has, more effort a bidder needs to put into the evaluation and preparation of the bids and hence the entry cost increases. It should be noted, however, because we do not directly observe the entry cost k in the data, we cannot conduct a direct test of Proposition 1 about the relationship between q^* and k .

4.2 The Relationship between b and N

Second, we empirically examine the relationship between the equilibrium bids and the number of potential bidders. We use $\log(\text{bids})$ as our dependent variable and the same set of exogenous variables as used in the regression in the last subsection.

We run two regressions. The first one is to pool all the bids together and use the standard linear OLS regression. The OLS regression method is easy to implement but does not address two important features of the auction data. First, we have 553 auctions with a total of 1606 bids. Most of the auctions have more than one bid. This data structure is more like an unbalanced panel data than a cross-section. Second, the OLS method does not control for the unobserved auction heterogeneity that is observed by bidders when making their decisions but unobserved by the econometrician. In view of these, besides the OLS, we estimate a random-effects panel data model.

The results from the two regressions are reported in Table 3 and Table 4. When we do not control for the unobserved heterogeneity, the OLS regression model fits the model well, with an adjusted R^2 to be 0.9324, indicating that we include in the regression the most important exogenous variables. In this regression, the number of potential bidders does not have a significant effect on bids at the 5% level with a t -statistic value of -1.80. We then estimate a random effects panel data model in order to control for the unobserved auction heterogeneity. The results are reported in Table 4 indicating that the number of potential bidders is even less significant than in the OLS estimation with a t -value of -1.14. Furthermore, the results from Table 4 support the existence of the unobserved auction heterogeneity as the error variance from the unobserved heterogeneity accounts for 36% of the total error variance. The statistical insignificance of N in explaining b can be viewed consistent with the implication from our model as given in Proposition 3. At the same time, it calls for a further structural analysis if one wants to quantify the “entry effect” and the “competition effect.”

4.3 The Relationship between Winning Bid and N

Finally, we empirically examine the relationship between the winning bid and the number of potential bidders. We use $\log(\text{winning bid})$ as our dependent variable and the same set of exogenous variables as used in the last two subsections.

The result from the OLS regression is reported in Table 5. In this regression, again, the number of potential bidders does not have a statistically significant effect on the winning bid. Furthermore, as we learn from the last subsection, ignoring the unobserved heterogeneity as in the OLS regression tends to over estimate the significance of this parameter. Therefore, this result can also be regarded as consistent with the implication from our model as given in Proposition 4. Again, it calls for a further structural analysis if one wants to quantify the “entry effect” and the “competition effect” of the number of potential bidders on the winning bid.

5 (Joint) Structural Inference of the Entry and Bidding Equilibrium Models

In this section, we derive structural models for entry and bidding from the game-theoretic models proposed in the last section. We then propose to use the semiparametric Bayesian method to

jointly estimate the structural model with the data from TDoT. The semiparametric Bayesian method enables us to jointly estimate the entry and bidding model with the unobserved auction heterogeneity being controlled for.

5.1 Structural Econometric Framework

In an econometric investigation, one often considers more than one auction, and the statistical inference is usually based on the assumption that the number of auctions approaches infinity. Therefore, possible heterogeneity across auctions needs to be taken into account. This issue can be addressed by modeling the distribution of the private costs to depend on the heterogeneity of auctioned objects and thus to vary across auctions. Specifically, let $F_\ell(\cdot)$ denote the distribution of private costs for the ℓ -th auction, $\ell = 1, \dots, L$, where L is the number of auctions. Assume that $F_\ell = F(\cdot|\mathbf{x}_\ell, u_\ell, \beta)$, where \mathbf{x}_ℓ is a vector of variables that represents the observed auction heterogeneity, and u_ℓ is a variable representing the unobserved auction heterogeneity, both affecting bidders' costs, and β is an unknown parameter vector in $B \subset \mathbb{R}^{m_1}$. u is assumed to be independent of \mathbf{x} . Let $f(\cdot|\mathbf{x}_\ell, u_\ell, \beta)$ denote the corresponding density for bidders' private costs. Let N_ℓ denote the number of potential bidders and n_ℓ denote the number of actual bidders at the ℓ -th auction. An econometric issue arising from modeling entry is that the entry cost k_ℓ at each auction is not observed. To resolve this, we assume that k_ℓ , $\ell = 1, \dots, L$, are randomly drawn from a distribution $H(\cdot|\mathbf{x}_\ell, u_\ell, \boldsymbol{\delta})$ with a density $h(\cdot|\mathbf{x}_\ell, u_\ell, \boldsymbol{\delta})$, where $\boldsymbol{\delta}$ is an unknown parameter vector in $\Psi \subset \mathbb{R}^{m_2}$.

5.2 Specification and Solving the Equilibrium Models

For a structural auction model with endogenous entry and unobserved heterogeneity, fully non-parametric estimation approach becomes infeasible and the classical likelihood or moment based estimation approach becomes extremely complex.¹⁵ In view of this, we propose to use the recently developed semiparametric Bayesian estimation method to estimate the entry and bidding models jointly. This approach offers several considerable advantages. First, with data augmentation, the

¹⁵For nonparametric inference of auction models, see Guerre, Perrigne and Vuong (2000), Li, Perrigne and Vuong (2000, 2002), Athey and Haile (2002), Haile, Hong and Shum (2002), Hendricks, Pinkse and Porter (2003) among others that are thoroughly surveyed in Athey and Haile (2004). For classical likelihood or moment based approaches, see Paarsch (1992), Donald and Paarsch (1993, 1996), Laffont, Ossard and Vuong (1995), Hong and Shum (2003), and Li (2005), to name a few.

Bayesian Markov Chain Monte Carlo (MCMC) estimation method is relatively straightforward to implement. Second, the MCMC gives us the finite-sample properties of the resulting estimates. Third, incorporating a nonparametric unobserved heterogeneity component makes the specification of the model more flexible and hence the results more robust. Fourth, because of the parameter-dependent support arising from structural auction models, classical likelihood based inference becomes non-standard and computationally intensive as well.

In this paper, we parameterize the density of contractors' private costs for completing the project as follows

$$f(c|\mathbf{x}, u) = \frac{1}{\exp(\alpha + \mathbf{x}\boldsymbol{\beta} + u)} \exp\left[-\frac{1}{\exp(\alpha + \mathbf{x}\boldsymbol{\beta} + u)}c\right], \quad (10)$$

for $c \in (0, \infty)$. The inclusion of the constant term α allows the normalization of u such that $E[u] = 0$. As a result, the equilibrium bidding strategy given in (6) becomes¹⁶:

$$s(c) = c + \frac{\sum_{j=2}^N P_B(n = j|n \geq 2) \exp\left[-\frac{j-1}{\exp(\mu_1)}c\right] \frac{\exp(\mu_1)}{j-1}}{\sum_{j=2}^N P_B(n = j|n \geq 2) \exp\left[-\frac{j-1}{\exp(\mu_1)}c\right]} \quad (11)$$

where $\mu_1 = \alpha + \mathbf{x}\boldsymbol{\beta} + u$. We further specify the entry cost density as follows:

$$h(k|\mathbf{x}, u) = \frac{1}{\exp(\gamma + \mathbf{x}\boldsymbol{\delta} + u)} \exp\left[-\frac{1}{\exp(\gamma + \mathbf{x}\boldsymbol{\delta} + u)}k\right], \quad (12)$$

for $k \in (0, \infty)$.

Note that while we employ the exponential distributions for specifying the conditional distributions of the private cost and entry cost given both observed and unobserved heterogeneities \mathbf{x} and u , we leave the distribution of the unobserved heterogeneity u unspecified and will use the data to reveal the distribution of u .¹⁷ Therefore, our specification and the resulting approach are

¹⁶See the derivation in Appendix.

¹⁷The literature on parametric estimation of auction models has usually used extreme value distributions such as Weibull and exponential to specify private values distributions. See, e.g., Paarsch (1992, 1997), Donald and Paarsch (1993, 1996), Haile (2001), Athey, Levin and Seira (2004). The specification of the exponential distributions for the conditional distributions of the private cost and the entry cost given both observed and unobserved heterogeneities is not as restrictive as it seems. This is because while this specification imposes that the conditional mean and the conditional variance of the private cost (or the entry cost) are the same given the observed and unobserved heterogeneities, this equality no longer holds after the unobserved heterogeneity is integrated out. The same approach has been widely used in microeconometrics such as in modeling counts or durations while controlling for unobserved heterogeneity. Furthermore, in our setup, the distribution of the unobserved auction heterogeneity is unspecified, making our approach more robust. For modeling unobserved heterogeneity in micro data, see Heckman (2001) for an insightful discussion.

semiparametric in nature, and are expected to yield more robust results than a fully parametric specification. Of course, these specifications help achieve identification of the structural parameters and the distribution of the unobserved auction heterogeneity from a classical viewpoint.

For the entry equilibrium, under these specifications for conditional densities of private costs and entry costs, entry equilibrium equation (7) can be written as¹⁸

$$\sum_{j=2}^N \left\{ \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{1 - (1-q^*)^{N-1}} \frac{\exp(\mu_1)}{j(j-1)} \right\} = k. \quad (13)$$

Note that the term

$$\sum_{j=2}^N \left\{ \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{1 - (1-q^*)^{N-1}} \frac{1}{j(j-1)} \right\} \quad (14)$$

is a function of q^* only given N and it goes to its maximum value 0.5 as $q^* \rightarrow 0$. Therefore, the entry cost equation implies that for this model, the entry costs in the observed auctions must satisfy the following restriction $k \leq 0.5 \exp(\mu_1)$, which has to be taken into account in estimation.

5.3 Semiparametric Bayesian Joint Estimation of Entry and Bidding

Our objective is to estimate the entry and bidding models jointly using both the observed numbers of actual bidders and bids. Joint estimation of the entry and bidding models is intractable if not impossible using the classical likelihood or moment based methods because of the complexity of the two models and the presence of latent variables in both models. The Bayesian method, on the other hand, provides a computationally feasible alternative because of the use of the data augmentation and MCMC techniques. To make use of the semiparametric Bayesian estimation and the MCMC algorithm, we first make some transformations. Let $e_{new} = \alpha + u$ represent the (unnormalized) unobserved heterogeneity term. Then $\mu_1 = \mathbf{x}\beta + e_{new}$. With the private cost distribution specified

¹⁸See the derivation in Appendix.

above and equation (11) we can obtain the following implied distributions for the equilibrium bids:

$$f(b|\mu_1) = \frac{1}{\exp(\mu_1)} \exp\left[-\frac{1}{\exp(\mu_1)} \varphi(b)\right] \left| \frac{\partial \varphi(b)}{\partial b} \right|, \quad (15)$$

for $b \in [\sum_{j=2}^N P_B(n=j|n \geq 2) \frac{\exp(\mu_1)}{j-1} / \sum_{j=2}^N P_B(n=j|n \geq 2), \infty)$, where $\varphi(b)$ is the inverse bidding function, and

$$\frac{\partial \varphi(b)}{\partial b} = \frac{\left[\sum_{j=2}^N P_B(n=j|n \geq 2) \exp\left(-\frac{j-1}{\exp(\mu_1)} \varphi(b)\right) \right]^2}{\left[\sum_{j=2}^N P_B(n=j|n \geq 2) \exp\left(-\frac{j-1}{\exp(\mu_1)} \varphi(b)\right) \frac{1}{j-1} \right] \left[\sum_{j=2}^N P_B(n=j|n \geq 2) \exp\left(-\frac{j-1}{\exp(\mu_1)} \varphi(b)\right) (j-1) \right]}. \quad (16)$$

Note that the lower support of b expressed here clearly indicates its dependence on the structural parameters, which violates the regularity conditions of the classical maximum likelihood estimation (Donald and Paarsch (1993, 1996), Chernozhukov and Hong (2004), Hirano and Porter (2003)). Also note that although equation (11) gives a complicated relationship between b and c , this relationship is monotone increasing. Moreover, it is computationally tractable to compute both $\varphi(b)$ and $\partial \varphi(b)/\partial b$, and hence to evaluate (15). Also, define

$$\mu_2 = \mu_1 - \mathbf{x} \boldsymbol{\delta}^* - \gamma^*$$

where $\boldsymbol{\delta}^* = \beta - \boldsymbol{\delta}$, $\gamma^* = \alpha - \gamma$. we can rewrite (12) as

$$h(k|\mu_1, \boldsymbol{\delta}^*, \gamma^*) = \frac{1}{\exp(\mu_1 - \mathbf{x} \boldsymbol{\delta}^* - \gamma^*)} \exp\left[-\frac{1}{\exp(\mu_1 - \mathbf{x} \boldsymbol{\delta}^* - \gamma^*)} k\right], \quad (17)$$

for $k \in (0, \infty)$.

In this mixed-strategy entry model, using (13) and (17), because of the one-to-one (inverse) relationship between q^* and k as discussed previously, the density for q^* implied by the density of k is

$$\begin{aligned} p(q^*|\mu_1, \boldsymbol{\delta}^*, \gamma^*) \\ = h(k|\mu_1, \boldsymbol{\delta}^*, \gamma^*) \times \left| \frac{\partial k}{\partial q^*} \right| \times \mathbb{I}(k \leq 0.5 \exp(\mu_1)), \end{aligned} \quad (18)$$

where $\mathbb{I}(\cdot)$ is an indicator function and

$$\frac{\partial k}{\partial q^*} = \sum_{j=2}^N \left\{ \binom{N-1}{j-1} \frac{\exp(\mu_1)}{j(j-1)} \left[\frac{(j-1)(q^*)^{j-2}(1-q^*)^{N-j}}{1-(1-q^*)^{N-1}} + \frac{(N-j)(q^*)^{j-1}(1-q^*)^{N-j-1}}{1-(1-q^*)^{N-1}} - \frac{(N-1)(q^*)^{j-1}(1-q^*)^{2N-j-2}}{(1-(1-q^*)^{N-1})^2} \right] \right\}. \quad (19)$$

Since the distributions of u and thus of $e_{new} = \mu_1 - \mathbf{x} \beta$ are left unspecified, we use a nonparametric method to approximate it. Specifically, we use an infinite mixture of normals to approximate the

unknown distributions. This is justified because Ferguson (1983) notes that any probability density function can be approximated arbitrarily closely in the L_1 norm by a countable mixture of normal densities.

$$g(\cdot) = \sum_{j=1}^{\infty} p_j \phi(\cdot | d_j, \sigma_j^2), \quad (20)$$

where $p_j \geq 0$, $\sum_{j=1}^{\infty} p_j = 1$ and $\phi(\cdot | d_j, \sigma_j^2)$ denotes the probability density function for a normal distribution with mean d_j and variance σ_j^2 . Note that this is a different specification from the finite mixture of normals as used in Geweke and Keane (2001) since we do not have a prior number of components in the mixture of normals. Instead, we will use the Dirichlet process prior to carry out the Bayesian nonparametric density estimation and update the number of components in the infinite mixture of normals and the mean and variance for each component.

We use the data augmentation approach (Tanner and Wong (1987)) and include unobservables $\mu_{1,\ell}$ and q_ℓ^* in the algorithm drawing them at each iteration and for all observations. Our semiparametric Bayesian estimation consists of two main parts. At each iteration, in the first part, using the augmented latent variables $\mu_{1,\ell}$ and q_ℓ^* and the current value of parameters, we can recover the unobserved heterogeneity terms through the relationship $e_{new,\ell} = \mu_{1,\ell} - \mathbf{x}_\ell \beta$. After recovering the unobserved heterogeneity terms, we can use a Bayesian approach explained in detail in Appendix to estimate their densities with a Dirichlet process prior for the unknown densities. We update the number of components (denoted by m_c , say) in approximating mixture of normals and the mean and variance of the normal denoted by d_ℓ and σ_ℓ^2 respectively for each ℓ . This approach was introduced by Lo (1984) and Ferguson (1983), with later work by Escobar (1994), Escobar and West (1995), and West, Müller, and Escobar (1994) discussing its computational issues.

In the second part of each iteration, we update the model parameters and values for the latent variables. Denote $\Theta = (\boldsymbol{\delta}^*, \gamma^*)$. We will use the following prior distributions: $\beta \sim Normal(\beta_0, B_0^{-1})$, $\Theta \sim Normal(\theta_0, D_0^{-1})$, where β_0 , B_0^{-1} , θ_0 , D_0^{-1} are known parameters. $Normal(\beta_0, B_0^{-1})$ denotes a multivariate normal distribution with mean vector β_0 and covariance matrix B_0^{-1} . Denote $\Delta = (\beta, \Theta)$ and $\mathbf{x}_\ell^* = (1, \mathbf{x}_\ell)$. Then the joint posterior density of the parameters and unobservables $\mu_{1,\ell}$ and

q_ℓ^* given the data and the distribution of the unobserved heterogeneity terms is

$$\begin{aligned}
& \pi(\Delta, \{\mu_{1,\ell}, q_\ell^*\}_{\ell=1}^L | b, n, \{d_\ell, \sigma_\ell^2\}_{\ell=1}^L) \\
& \propto \text{prior}(\Delta) \times \prod_{\ell=1}^L p(b_{1\ell}, \dots, b_{n_\ell\ell} | \mu_{1,\ell}, q_\ell^*) \times p(n_\ell | n_\ell \geq 2, q_\ell^*) \times p(q_\ell^* | \mu_{1,\ell}, \Delta) \\
& \quad \times p(\mu_{1,\ell} | d_\ell, \sigma_\ell^2) \prod_{i=1}^{n_\ell} \mathbb{I} \left[b_{i\ell} \geq \sum_{j=2}^{N_\ell} P_B(n_\ell = j | n_\ell \geq 2) \frac{\exp(\mu_{1,\ell})}{(j-1)} \right] \\
& \propto \text{prior}(\Delta) \\
& \quad \times \prod_{\ell=1}^L \left\{ \frac{1}{\exp(n_\ell \mu_{1,\ell})} \left\{ -\frac{1}{\exp(\mu_{1,\ell})} \sum_{i=1}^{n_\ell} \varphi(b_{i\ell}) \right\} \prod_{i=1}^{n_\ell} \left| \frac{\partial \varphi(b_{i\ell})}{\partial b_{i\ell}} \right| \right. \\
& \quad \times \prod_{i=1}^{n_\ell} \mathbb{I} \left[b_{i\ell} \geq \sum_{j=2}^{N_\ell} P_B(n_\ell = j | n_\ell \geq 2) \frac{\exp(\mu_{1,\ell})}{(j-1)} \right] \\
& \quad \times \frac{\binom{N_\ell}{n_\ell} (q_\ell^*)^{n_\ell} (1 - q_\ell^*)^{N_\ell - n_\ell}}{1 - (1 - q_\ell^*)^{N_\ell} - N_\ell q_\ell^* (1 - q_\ell^*)^{N_\ell - 1}} \\
& \quad \times \frac{1}{\exp(\mu_\ell - \mathbf{x}_\ell^* \Theta)} \exp\left(-\frac{1}{\exp(\mu_{1,\ell} - \mathbf{x}_\ell^* \Theta)} k_\ell\right) \\
& \quad \times \left| \frac{\partial k_\ell}{\partial q_\ell^*} \right| \times 1(k_\ell \leq 0.5 \exp(\mu_{1,\ell})) \\
& \quad \times \frac{1}{\sqrt{2\pi\sigma_\ell^2}} \exp[-0.5\sigma_\ell^{-2}(\mu_{1,\ell} - \mathbf{x}_\ell \boldsymbol{\beta} - d_\ell)^2] \tag{21}
\end{aligned}$$

where¹⁹

$$k_\ell = \sum_{j=2}^{N_\ell} \left\{ \frac{\binom{N_\ell - 1}{j-1} (q_\ell^*)^{j-1} (1 - q_\ell^*)^{N_\ell - j} \exp(\mu_{1,\ell})}{1 - (1 - q_\ell^*)^{N_\ell - 1} j(j-1)} \right\}. \tag{22}$$

We construct our Markov chain blocking the parameters and the latent variables as $(\mu_{1,\ell}, q_\ell^*)$, $\boldsymbol{\beta}$, Θ and $(d_\ell, \sigma_\ell^2, m_c)$ with the full conditional distributions: $[\mu_{1,\ell}, q_\ell^* | \boldsymbol{\beta}, \boldsymbol{\delta}^*, \gamma^* d_\ell, \sigma_\ell^2]$, $[\boldsymbol{\beta} | \mu_{1,\ell}, d_\ell, \sigma_\ell^2]$,

¹⁹Note that here we modify the conditional distribution of the number of actual bidders to be a truncated binomial distribution from below at 2. This is because we concentrate on the subsample of auctions with at least two actual bidders in the structural estimation. This subset consists of 97.5% of the data; by the assumption of our model the government does not play any role in these auctions.

$[\Theta|\mu_{1,\ell}, q_\ell^*], [d_\ell, \sigma_\ell^2, m_c|\mu_{1,\ell}, \beta]$. The following steps summarize our algorithm.²⁰

Sampling $(\mu_{1,\ell}, q_\ell^*)$. Since the evaluation of the likelihood involves computing the inverse bidding function $\phi(b_{i\ell})$ at each iteration and for each bid, it is quite time consuming. In order to save computational time, we propose to block $(\mu_{1,\ell}, q_\ell^*)$ together rather than update them individually. Note also that the posterior density for $(\mu_{1,\ell}, q_\ell^*)$ does not have a form that can facilitate a direct random draw from it. To overcome the associated computational problem, we propose to utilize the Metropolis-Hasting (MH) algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953), Hastings (1970)) to draw from the density.

Sampling β . Draw β given $\mu_{1,\ell}, d_\ell, \sigma_\ell^2$ and it's prior, which is a normal distribution with variance $\Lambda = \left(B_0 + \sum_{\ell=1}^L \sigma_\ell^{-2} \mathbf{x}'_\ell \mathbf{x}_\ell\right)^{-1}$ and mean $\bar{\beta} = \Lambda \left(B_0 \beta_0 + \sum_{\ell=1}^L \sigma_\ell^{-2} \mathbf{x}'_\ell (\mu_{1,\ell} - d_\ell)\right)$.

Sampling $\Theta = (\delta^*, \gamma^*)$. The full conditional density for $\Theta = (\delta^*, \gamma^*)$ is

$$\begin{aligned} \pi[\Theta|\mu_{1,\ell}, q_\ell^*] &= \exp\left[-(\Theta - \theta_0)' D_0 (\Theta - \theta_0) / 2\right] \\ &\times \prod_{\ell=1}^L \frac{1}{\exp(\mu_{1,\ell} - \mathbf{x}'_\ell \Theta)} \exp\left(-\frac{1}{\exp(\mu_{1,\ell} - \mathbf{x}'_\ell \Theta)} k_\ell\right) \end{aligned} \quad (23)$$

This posterior density does not allow one to make direct random draws. Again, we utilize the MH algorithm.

Updating d_ℓ, σ_ℓ^2 and m_c . Get $e_{new,\ell} = \mu_{1,\ell} - \mathbf{x}_\ell \beta$ and update d_ℓ, σ_ℓ^2 , and m_c using the non-parametric Bayesian estimation method as outlined in Appendix.

These steps constitute one MCMC iteration. In the end, we will be able to obtain the posterior densities for β, δ , and $\gamma - \alpha$, as well as the nonparametric density for e_{new} .²¹

6 Results

In this section, we present the estimation results for the structural elements of the joint bidding and entry model. Since the mowing jobs are relatively simple, our data provide us enough information to control for the observed auction heterogeneity. We choose $\mathbf{x} = \{\log(\text{Estimate}), \log(\text{Day}), \log(\text{Full}), \log(\text{Other}), \text{Items}, \text{State}, \text{Interstate}\}$. See the summary statistics and variable definitions in Table 1. We center priors for parameters $\beta \sim N(\mathbf{0}_7, 10I_7)$, $\Theta \sim N(\mathbf{0}_8, 10I_8)$ to reflect weak prior information

²⁰See Appendix for details.

²¹Since $e_{new} = \alpha + u$ with $E[u] = 0$, we can also recover α from the estimated $E[e_{new}]$, and thus γ .

on these parameters. $\mathbf{0}_7$ denotes a vector of length 7 for 0's and I_7 is the 7-dimensional identity matrix. In the nonparametric component, we set $\mu_0 = 0$, $\tau_0 = 4$, $N_0 = 5$, $R_0 = 10$ and $d_1 = d_2 = 2$ since there is no information on the unobserved heterogeneity. Finally, the tuning parameters are chosen to be $h_0 = 0.5$, $v_0 = v_1 = 15$ and $\tau_2 = 1$ to obtain reasonable acceptance rates in the MH steps within the Gibbs Sampler.

We run the MCMC for 15,000 iterations, with an initial 5,000 burn-in period. Figure (2) and Figure (3) provide the time series plots, the histograms and the autocorregrams for the sample draws. This MCMC sample size design appears to be adequate since the autocorrelations in the figures indicate that our sample mixes relatively well. With the current setting of tuning parameters, the average acceptance rates are 12.12% for sampling $(\mu_{1,\ell}, q_\ell^*)$ and 34.03% for sampling Θ . The computation time for this algorithm is about 12.64 seconds per draw on average on a Pentium® 4 2.40 GHz processor. More than half of the computation time is from calculating the inverse bidding function. This computational burden is bearable considering the complexity of the estimation problem we try to solve.

6.1 Parameter Estimates

We summarize the estimation results from the semiparametric Bayesian algorithm in Table 6. Note that in the private value distribution, only the variable $\log(\text{Estimate})$ has a large positive effect on the mean of bidders' private cost for completing the job. The estimated coefficient is 0.9425, which indicates that the engineer's estimate is a good estimate for the costs of completing the job and bidders use this information to determine their own private costs for a contract. 1 unit increase in the $\log(\text{Estimate})$ will increase the mean of bidders' private cost by 1.57 times.

Regarding the entry cost distribution, we also have interesting results. First, again, the $\log(\text{Estimate})$ turns out to have a large positive effect on the entry cost. This is reasonable since contracts with higher values usually involve more tasks and hence more cost estimation work and more paper work for preparing the bids. Second, the estimate for the variable Items is 0.1817. This tells us that when the auction has one more item, the entry cost is going to increase for about 19.93%. Also, a state job will increase the entry cost for about 23.53%. This may be due to the fact that the state and local jobs have different requirements for preparing for the bids. Also, the number of working days in the job and whether the job is on an interstate highway appear to have small positive effects on

the entry cost. Other variables seem to have no significant effects.

Furthermore, we use the structural estimates to find the entry cost in relation to the private cost by simulating the entry cost and the private cost from the corresponding estimated distributions for each auction. The average ratio between the entry cost and the private cost is about 13.8%, meaning that the entry cost accounts for about 13.8% of the private cost. This finding indicates that the entry cost is a significant part of bidders' decision making process and hence a theoretical model or an empirical analysis that ignore the entry effect may lead to misleading policy recommendations.

We have also used the structural estimates to simulate the winner's payoff defined as the winning bid less the sum of the private cost and the entry cost, as well as the winner's information rent defined as the ratio between the winner's payoff and the winning bid. It turns out that the average winner's payoff is \$40,020 and the average information rent is 0.26.

6.2 Density Estimate for the Unobserved Heterogeneity

In the Bayesian framework, an informative way to draw implications from the unknown densities of the error term $e_{new,\ell}$ is to study its predictive distribution since it summarizes the information on the data. Parallel to equation (A.18), conditional on $(\theta_1, \dots, \theta_L)$, for the new unit $\ell = L + 1$, we have

$$\theta_{L+1} | (\theta_1, \dots, \theta_L) \sim \frac{\tau}{\tau + L} P_0 + \frac{1}{\tau + L} \sum_{j=1}^{m_c} n_j 1(\xi_j), \quad (24)$$

where n_j is the number of θ_i 's taking the value ξ_j . Thus, the distribution of $e_{new,\ell+1}$ conditional on the data can be rewritten as

$$\begin{aligned} q(e_{new,L+1} | \theta_1, \dots, \theta_L) &= \frac{\tau}{\tau + L} qt(e_{new,L+1} | \mu_0, (1 + \tau_0)R_0/n_0, n_0) \\ &\quad + \frac{1}{\tau + L} \sum_{j=1}^{m_c} n_j \phi(e_{new,L+1} | \xi_j). \end{aligned} \quad (25)$$

Thus the predictive distribution for the error term $e_{new,L+1}$ can be obtained as

$$q(e_{new,L+1} | data) = \int q(e_{new,L+1} | \theta_1, \dots, \theta_L) \pi(\theta_1, \dots, \theta_L | data) d(\theta_1, \dots, \theta_L). \quad (26)$$

Since the Gibbs sampler provides draws for θ_ℓ 's, we can use the Monte Carlo method to integrate out θ_ℓ 's to estimate $q(e_{new,L+1} | data)$ as

$$\hat{q}(e_{new,L+1} | data) = \frac{1}{M} \sum_{i=1}^M q(e_{new,L+1} | \theta_1^{(i)}, \dots, \theta_L^{(i)}), \quad (27)$$

where $(\theta_1^{(i)}, \dots, \theta_L^{(i)})$ is a simulated sample of $(\theta_1, \dots, \theta_L)$.

Figure (4) plots the estimated predictive density for the unobserved term e_{new} . This indicates that there exists substantial unobserved heterogeneity in both the private value and entry cost distributions. The mode of the density is around -0.3. Furthermore, notice that the estimated density is asymmetric and slightly left skewed. This is an evidence of the nonnormality of the unobserved heterogeneity, which demonstrates the need for the nonparametric modelling of this density.

6.3 Model Fit

To assess the fit of our model, we simulate auctions in our data set using the structural estimates we obtain from the semiparametric Bayesian estimation. For each auction ℓ in the data, we simulate an entry cost k from the exponential density with mean $\exp(\overline{\mu_{2,\ell}})$, where $\overline{\mu_{2,\ell}}$ is the posterior mean of the draws of $\mu_{2,\ell}$ during the MCMC iterations. The drawn k must satisfy the restriction required by the model, that is, $k \leq 0.5 \exp(\overline{\mu_{1,\ell}})$. We then use this entry cost, together with $\exp(\overline{\mu_{1,\ell}})$, which is the posterior mean of the draws of $\exp(\mu_{1,\ell})$ during the MCMC iterations and other auction specific covariates to calculate the equilibrium entry probability for this auction. We then simulate the number of the actual bidders n using the equilibrium entry probability and the number of potential bidders for this auction. Then, we simulate n private costs from the exponential distribution with the mean of $\exp(\overline{\mu_{1,\ell}})$. Finally, we calculate the equilibrium bids using the equilibrium bidding functions and determine the procurement cost for this auction. We repeat this procedure for 500 times to obtain a simulated mean of the procurement cost for this auction. We perform this analysis for all the auctions in our data and compare the simulated procurement costs with the actual procurement costs in our data.

Figure (5) presents this comparison. The upper panel displays the histogram for the actual procurement costs in our data while the lower panel displays the histogram for the means of the simulated procurement costs. As one can see, the two histograms are very similar to each other, indicating that our model fits the data very well.

7 Counterfactual Analysis

7.1 Quantifying the “Competition Effect” and the “Entry Effect” of Number of Potential Bidders N on b

One interesting theoretical relationship we obtain from Section 3 is that in our model, the relationship between the number of potential bidders, N , and the individual bid, b , may not be monotone decreasing. This is due to the two effects of N on b . One is the “competition effect” and the other is the “entry effect.” As given in Proposition 3, the “competition effect” is always negative and the “entry effect” is always positive. As a result, as the number of potential bidders increases, the equilibrium bid is increasing as well, provided that the positive “entry effect” dominates the negative “competition effect.” If this is the case, policies that encourage more potential bidders might actually increase the equilibrium bid. One advantage of the structural approach is that it enables us to use the structural estimates to conduct a counterfactual analysis to quantify the “competition effect” and the “entry effect” of N on b separately.

To quantify the “competition effect” and the “entry effect,” we run a counterfactual experiment for a representative auction. We pick the 123th auction in our dataset. This auction has 11 potential bidders, which is the mean value for the number of potential bidders in our dataset. Also, the engineer’s estimate for this auction is \$170,060, which is very close to the mean value of engineer’s estimate \$165,348.90. We use the last 1,000 iterations of our MCMC algorithm for this counterfactual experiment. At each iteration, with the simulated values of $\mu_{1,123}$ and k_{123} , where the subscript 123 denotes the 123th auction, we can recover a unique q_{123}^* for this auction for each value of N . To quantify the “competition effect,” we set the private cost c at $\exp(\overline{\mu_{1,123}})$, where $\overline{\mu_{1,123}}$ is the posterior mean of the draws of $\mu_{1,123}$ during the MCMC iterations across all the iterations and then calculate the equilibrium bid using the equilibrium bidding function. We vary the number of potential bidders from 3 to 26, which are the minimum and the maximum observed values in the data. Note that as we vary the number of potential bidders from N to $N + 1$, the q_{123}^* is fixed. For each value of the potential bidders N , we report the change in the equilibrium bid from N to $N + 1$ as the “competition effect.”

On the other hand, quantifying the “entry effect” directly is difficult, but quantifying the “total effect” from N to b is relatively straightforward. After we get the “total effect,” the “entry effect”

can be easily computed as the difference between the “total effect” and the “competition effect.” To calculate the “total effect,” we use almost the same simulation setup as in quantifying the “competition effect.” The only difference here is that as we vary the number of potential bidders from N to $N + 1$, we obtain a new q_{123}^* instead of using the fixed q_{123}^* . For each value of the potential bidders N , we report the change in the equilibrium bid from this simulation as the “total effect.”

We plot the simulated mean of the “competition effect,” the “entry effect” and the “total effect” and their corresponding confidence bands from the counterfactual experiment in Figure (6). The x -axis denotes the number of potential bidders N and the y -axis represents the change in the equilibrium bid from increasing the number of potential bidders from N to $N + 1$. Figure (6) reveals several interesting results. First, an obvious feature of the plot is that when the number of potential bidders increases, the “competition effect” is always negative. This is consistent with the theory since as the number of potential bidders increases and the equilibrium entry probability is fixed, then bidders would anticipate more actual bidders and bid more aggressively. Second, the “entry effect” is always positive, which is also consistent with what our model implies. This shows that the entry behavior in these auctions actually make the bidders to bid less aggressively, all else being equal. Third, the simulated mean of “total effect” from N on b is positive for $N \geq 7$. This demonstrates that in our simulation setup, on average, the positive “entry effect” weakly dominates the negative “competition effect” for a wide range of the number of potential bidders. In more detail, when $N = 7$, the simulated mean of the equilibrium bid for this representative auction is \$146,431.45. But the simulated mean of the equilibrium bid becomes \$149,704.62 (or 1.83% increase) when $N = 12$ and \$154,328.57 (or 5.39% increase) when $N = 26$. This finding is interesting since in the usual IPV auction model without entry, bidders will bid more aggressively when facing a larger number of potential bidders. In our case, however, when the endogenous entry matters and the bidders do not know the number of actual bidders when they submit bids, the bids actually become less aggressive.

7.2 Quantifying the “Competition Effect” and the “Entry Effect” of Number of Potential Bidders N on Procurement Costs

Another interesting theoretical relationship we obtain from Section 3 is that in our model, the relationship between the number of potential bidders, N , and the procurement cost, W , may not

be monotone decreasing. Again, this is due to the two effects of N on W . One is the “competition effect” and the other is the “entry effect.” As given in Proposition 4, the “competition effect” can be negative and the “entry effect” can be positive. If this is indeed the case, then as the number of potential bidders increases, the government’s procurement cost is increasing as well, provided that the positive “entry effect” dominates the negative “competition effect.” This means that policies that encourage more potential bidders might actually increase the government’s procurement costs. We now use the structural estimates to conduct a counterfactual analysis to quantify the “competition effect” and the “entry effect” of N on the procurement costs separately.

To quantify the “competition effect” and the “entry effect,” we again run a counterfactual experiment for the same representative auction. As in the previous subsection, we use the last 1,000 iterations of our MCMC algorithm for this counterfactual experiment. At each iteration, with the simulated values of $\mu_{1,123}$ and k_{123} , we can recover a unique q_{123}^* for this auction for each value of N . To quantify the “competition effect,” we compute the expected winning bid using its formula derived in Lemma 5 in Appendix. We vary the number of potential bidders from 3 to 26. Note that as we vary the number of potential bidders from N to $N + 1$, the q_{123}^* is fixed. For each value of the potential bidders N , we report the change in the expected winning bid from N to $N + 1$ as the “competition effect.”

On the other hand, since quantifying the “entry effect” directly is difficult, we again first quantify the “total effect” from N to the procurement cost, which is relatively straightforward. After we get the “total effect,” the “entry effect” can be easily computed as the difference between the “total effect” and the “competition effect.” To calculate the “total effect,” we use almost the same setup as in quantifying the “competition effect.” The only difference here is that as we vary the number of potential bidders from N to $N + 1$, we obtain a new q_{123}^* instead of using the fixed q_{123}^* . For each value of the potential bidders N , we report the change in the expected winning bid as the “total effect.”

We plot the simulated mean of the “competition effect,” the “entry effect” and the “total effect” and their corresponding confidence bands from the simulation in Figure (7). The x -axis denotes the number of potential bidders N and the y -axis represents the change in the expected winning bid (procurement cost) from increasing the number of potential bidders from N to $N + 1$. Figure (7) reveals several interesting results. First, an obvious feature of the plot is that when the number of

potential bidders increases, the “competition effect” is always negative. Second, the “entry effect” is always positive. Third, the positive “entry effect” significantly dominates the negative “competition effect,” leading to a positive “total effect” for a wide range of the number of potential bidders. This implies that the relationship between N and W is not monotone decreasing, which is consistent with our theoretical result. We also find from the experiment that the expected winning bid hits its lowest point when $N = 3$, which is \$112,390.01 for this representative auction. But the expected winning bid becomes \$152,666.15 (or 35.84% increase) when $N = 12$ and \$152,165.71 (or 35.39% increase) when $N = 26$. This interesting finding indicates that with the endogenous entry and uncertain number of actual bidders, the procurement cost can actually become less aggressive, as the number of potential bidders increases. Thus, in this kind of auction environment, policies that encourage more potential bidders may not be desirable, especially when the existing number of potential bidders is small, as evidenced from our analysis.

7.3 Quantifying the Savings on Procurement Costs by Reducing the Entry Cost

From the estimation results, we find strong evidence that part of the bids are used to recover bidders’ entry cost. Therefore, it is interesting to quantify the savings on the government’s procurement costs by reducing bidders’ entry cost. This has important policy implications since if the savings is substantial, then the government can improve social welfare by designing mechanisms to reduce the entry cost.

To quantify the savings, we again run a counterfactual experiment for the same representative auction. We first set the entry cost k at $\exp(\overline{\mu_{2,123}})$, where $\overline{\mu_{2,123}}$ is the posterior mean of the draws of $\mu_{2,123}$ during the MCMC iterations across all the iterations and $u_{1,123}$ set at similarly defined $\overline{\mu_{1,123}}$. This allows us to recover a unique q_{123}^* for this auction and thus compute the expected winning bid using its formula derived in Lemma 5 in Appendix. We then re-do the analysis by setting the entry cost k at 20%, 50% and 80% of its original level, that is, at $0.2 * \exp(\overline{\mu_{2,123}})$, $0.5 * \exp(\overline{\mu_{2,123}})$ and $0.8 * \exp(\overline{\mu_{2,123}})$ respectively. This difference between the two settings can be regarded as the savings on procurement costs by reducing the entry cost. Our results show that by reducing the entry cost to half of its original level, the government can save approximately 28.18% of the procurement cost for this particular auction. The savings are 44.49% and 7.19% of the procurement cost if the entry cost is reduced to be 20% and 80% of its original level respectively.

8 Conclusion

In this paper, we propose a two-stage procurement auction model with endogenous entry and uncertain number of actual bidders. Our model yields several interesting implications, which are supported by the data we analyze. Most importantly, for the first time, we show that even within the IPV paradigm, as the number of potential bidders increases, bidders' equilibrium bidding behavior can become less aggressive. We also show that this can be the case for the procurement cost as well. Thus, increasing potential competition may not necessarily benefit the auctioneer. Whether this is the case is an empirical question that can be answered through a rigorous structural analysis, as evidenced from our paper.

In order to answer this question, among others that may be of interest to the auctioneer or/and policy makers, we develop a fully structural model for entry and bidding based on the game-theoretic model of entry and bidding with uncertain number of actual bidders proposed in this paper. To circumvent the complexity associated with our structural model for entry and bidding and the need for controlling for the unobserved auction heterogeneity, we propose a new semiparametric Bayesian estimation algorithm to estimate the structural elements of the model. With our structural approach, we are able to signify the effect from the unobserved auction heterogeneity and quantify the “entry effect” and the “competition effect.” Therefore, our paper contributes to the literature by providing a unified framework to study procurement auctions with entry and unobserved auction heterogeneity, and sheds light on empirical analyses of other auctions.

Appendix I: Proofs

1. **Proof of Proposition 1** Using the equilibrium bidding strategy (6), the entry equation (2) can be rewritten as

$$\int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N [P_B(n=j|n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx] f(c) dc = k. \quad (\text{A.1})$$

Denote $\int_{\underline{c}}^{\bar{c}} \int_c^{\bar{c}} (1-F(x))^{j-1} dx] f(c) dc = E\pi_j$, we have the following equation:

$$\sum_{j=2}^N \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{1 - (1-q^*)^{N-1}} E\pi_j - k = 0. \quad (\text{A.2})$$

The differential of (A.2) with respect to q^* and k can be written as

$$\begin{aligned} & \sum_{j=2}^N \frac{\binom{N-1}{j-1} (j-1) (q^*)^{j-2} (1-q^*)^{N-j}}{1 - (1-q^*)^{N-1}} E\pi_j dq^* \\ & - \sum_{j=2}^N \frac{\binom{N-1}{j-1} (N-j) (q^*)^{j-1} (1-q^*)^{N-j-1}}{1 - (1-q^*)^{N-1}} E\pi_j dq^* \\ & - \sum_{j=2}^N \frac{\binom{N-1}{j-1} (N-1) (q^*)^{j-1} (1-q^*)^{N-j} (1-q^*)^{N-2}}{[1 - (1-q^*)^{N-1}]^2} E\pi_j dq^* \\ & = dk, \end{aligned}$$

which can be further simplified as

$$\begin{aligned} & \sum_{j=2}^N \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{[1 - (1-q^*)^{N-1}] q^* (1-q^*)} \\ & \left[(j-1)(1-q^*) - (N-j)q^* - \frac{q^*(N-1)(1-q^*)^{N-1}}{1 - (1-q^*)^{N-1}} \right] E\pi_j dq^* \\ & = dk \end{aligned}$$

Now, change the variable using $l = j - 1$, we get the following:

$$\frac{1}{q^*(1-q^*)} \sum_{l=1}^{N-1} \frac{\binom{N-1}{l} (q^*)^l (1-q^*)^{N-1-l}}{[1-(1-q^*)^{N-1}]} \left[l - \frac{(N-1)q^*}{1-(1-q^*)^{N-1}} \right] E\pi_{l+1} dq^* = dk$$

Note, here

$$\sum_{l=1}^{N-1} \frac{\binom{N-1}{l} (q^*)^l (1-q^*)^{N-1-l}}{[1-(1-q^*)^{N-1}]} \left[l - \frac{(N-1)q^*}{1-(1-q^*)^{N-1}} \right] E\pi_{l+1} = Cov(l, E\pi_{l+1})$$

since $\frac{\binom{N-1}{l} (q^*)^l (1-q^*)^{N-1-l}}{[1-(1-q^*)^{N-1}]}$ is $P_B(l|l \geq 1)$ and $E(l|l \geq 1) = \frac{(N-1)q^*}{1-(1-q^*)^{N-1}}$ from Lemma 1 below. Therefore,

$$\frac{dq^*}{dk} = \frac{q^*(1-q^*)}{Cov(l, E\pi_{l+1})} < 0 \text{ if } Cov(l, E\pi_{l+1}) < 0.$$

From the definition for $E\pi_j$ above, we can easily see that $Cov(l, E\pi_{l+1}) < 0$. This completes the proof.

2. Lemma 1

$$E(l|l \geq 1) = \frac{(N-1)q^*}{1-(1-q^*)^{N-1}}$$

Proof

$$\begin{aligned} E(l|l \geq 1) &= \sum_{l=1}^{N-1} \frac{\binom{N-1}{l} (q^*)^l (1-q^*)^{N-1-l}}{[1-(1-q^*)^{N-1}]} l \\ &= (N-1)q^* \sum_{l=1}^{N-1} \frac{\binom{N-2}{l-1} (q^*)^{l-1} (1-q^*)^{N-2-(l-1)}}{[1-(1-q^*)^{N-1}]} \end{aligned}$$

$$\begin{aligned}
&= \frac{(N-1)q^*}{[1-(1-q^*)^{N-1}]} \sum_{l=1}^{N-1} \binom{N-2}{l-1} (q^*)^{l-1} (1-q^*)^{N-2-(l-1)} \\
&= \frac{(N-1)q^*}{[1-(1-q^*)^{N-1}]} \sum_{m=0}^{N-2} \binom{N-2}{m} (q^*)^m (1-q^*)^{N-2-m} \\
&= \frac{(N-1)q^*}{[1-(1-q^*)^{N-1}]} (q^* + 1 - q^*)^{N-2} \\
&= \frac{(N-1)q^*}{[1-(1-q^*)^{N-1}]}.
\end{aligned}$$

3. **Proof of Proposition 2** Rewrite the entry equation (2) or (A.1) as $k = g(N, q^*)$, then by total differentiation of k with respect to N , we get $\frac{dk}{dN} = \frac{\partial k}{\partial N} + \frac{\partial k}{\partial q^*} \frac{\partial q^*}{\partial N}$. Set $\frac{dk}{dN} = 0$ since both k and N are exogenous, it gives

$$\frac{\partial q^*}{\partial N} = -\frac{\partial k}{\partial N} * \frac{\partial q^*}{\partial k}.$$

Since $\frac{\partial q^*}{\partial k} < 0$ from the last proposition, $\frac{\partial q^*}{\partial N}$ has the same sign as $\frac{\partial k}{\partial N}$. With $\frac{\partial k}{\partial N} < 0$ proved by Lemma 2 below, this completes the proof.

4. **Lemma 2** $\frac{\partial k}{\partial N} < 0$.

Proof By definition of equation (A.1),

$$\int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N [P_B(n=j|N \geq n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx] f(c) dc = k_N, \quad (\text{A.3})$$

where k_N denotes the equilibrium entry cost with N potential bidders and entry probability q^* . Similarly,

$$\int_{\underline{c}}^{\bar{c}} \sum_{j=2}^{N+1} [P_B(n=j|N+1 \geq n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx] f(c) dc = k_{N+1}. \quad (\text{A.4})$$

Since

$$\begin{aligned}
1 &= \sum_{j=2}^N P_B(n=j|N \geq n \geq 2) = \sum_{j=2}^N \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{1-(1-q^*)^{N-1}} \\
&= \sum_{j=2}^{N+1} \frac{\binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j}}{1-(1-q^*)^N} = \sum_{j=2}^{N+1} P_B(n=j|N+1 \geq n \geq 2),
\end{aligned}$$

\Leftrightarrow

$$\frac{(q^*)^N}{1 - (1 - q^*)^N} = \sum_{j=2}^N \left[\frac{\binom{N-1}{j-1} (q^*)^{j-1} (1 - q^*)^{N-j}}{1 - (1 - q^*)^{N-1}} - \frac{\binom{N}{j-1} (q^*)^{j-1} (1 - q^*)^{N+1-j}}{1 - (1 - q^*)^N} \right]. \quad (\text{A.5})$$

Using equations (A.3) and (A.4),

$$\begin{aligned} k_N - k_{N+1} &= \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left[\frac{\binom{N-1}{j-1} (q^*)^{j-1} (1 - q^*)^{N-j}}{1 - (1 - q^*)^{N-1}} \int_c^{\bar{c}} (1 - F(x))^{j-1} dx \right] f(c) dc \\ &\quad - \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^{N+1} \left[\frac{\binom{N}{j-1} (q^*)^{j-1} (1 - q^*)^{N+1-j}}{1 - (1 - q^*)^N} \int_c^{\bar{c}} (1 - F(x))^{j-1} dx \right] f(c) dc \\ &= \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left[\frac{\binom{N-1}{j-1} (q^*)^{j-1} (1 - q^*)^{N-j}}{1 - (1 - q^*)^{N-1}} \int_c^{\bar{c}} (1 - F(x))^{j-1} dx \right] f(c) dc \\ &\quad - \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left[\frac{\binom{N}{j-1} (q^*)^{j-1} (1 - q^*)^{N+1-j}}{1 - (1 - q^*)^N} \int_c^{\bar{c}} (1 - F(x))^{j-1} dx \right] f(c) dc \\ &\quad - \int_{\underline{c}}^{\bar{c}} \frac{(q^*)^N}{1 - (1 - q^*)^N} \int_c^{\bar{c}} (1 - F(x))^N dx f(c) dc \quad (\text{using equation (A.5)}) \\ &= \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left[A_j \int_c^{\bar{c}} (1 - F(x))^{j-1} dx \right] f(c) dc - \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left[A_j \int_c^{\bar{c}} (1 - F(x))^N dx \right] f(c) dc \\ &= \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left\{ A_j * \int_c^{\bar{c}} [(1 - F(x))^{j-1} - (1 - F(x))^N] dx \right\} f(c) dc, \end{aligned}$$

where

$$A_j = \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{1 - (1-q^*)^{N-1}} - \frac{\binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j}}{1 - (1-q^*)^N}.$$

If for all $j = 1, 2, \dots, N$, $A_j > 0$, then $k_N - k_{N+1} > 0$ and hence $\frac{\partial k}{\partial N} < 0$ as desired. Otherwise, note that

$$A_j = \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} \left[\frac{1}{1 - (1-q^*)^{N-1}} - \frac{\frac{N}{N-j+1}}{\frac{1}{1-q} - (1-q)^{N-1}} \right].$$

As j increases, $\frac{\frac{N}{N-j+1}}{\frac{1}{1-q} - (1-q)^{N-1}}$ increases. Thus, there exists a j_0 such that for $j < j_0$, $A_j > 0$ and for $j \geq j_0$, $A_j < 0$. Therefore,

$$\begin{aligned} k_N - k_{N+1} &\geq \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^{j_0-1} \left\{ A_j * \int_c^{\bar{c}} [(1-F(x))^{j_0-1} - (1-F(x))^N] dx \right\} f(c) dc \\ &\quad + \int_{\underline{c}}^{\bar{c}} \sum_{j=j_0}^N \left\{ A_j * \int_c^{\bar{c}} [(1-F(x))^{j_0-1} - (1-F(x))^N] dx \right\} f(c) dc \\ &\geq \int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N \left\{ A_j * \int_c^{\bar{c}} [(1-F(x))^{j_0-1} - (1-F(x))^N] dx \right\} f(c) dc \\ &= \int_{\underline{c}}^{\bar{c}} \frac{(q^*)^N}{1 - (1-q^*)^N} \int_c^{\bar{c}} [(1-F(x))^{j_0-1} - (1-F(x))^N] dx f(c) dc > 0, \end{aligned}$$

where we use the relationship $\frac{(q^*)^N}{1 - (1-q^*)^N} = \sum_{j=2}^N A_j$ from equation (A.5). This completes the proof.

5. **Proof of Proposition 3** Rewrite the equilibrium bidding strategy equation (6) as $b = p(N, q^*)$, the total differentiation of b with respect to N is

$$\frac{db}{dN} = \frac{\partial b}{\partial N} + \frac{\partial b}{\partial q^*} * \frac{\partial q^*}{\partial N}.$$

This equation illustrates that the number of potential bidders, N , affects the equilibrium bids in two channels. The first channel is the usual “competition effect,” as represented by $\frac{\partial b}{\partial N}$, which is negative as proved by Lemma 3 below. The second channel, which we call the “entry effect,” directs the effects of N on b through the entry stage. This “entry effect” is positive since $\frac{\partial b}{\partial q^*} < 0$ as proved in lemma 4 below and $\frac{\partial q^*}{\partial N} < 0$ as proved in proposition 2. Therefore,

the sign of $\frac{db}{dN}$ depends on which of the two effects, the “competition effect” or the “entry effect,” is large. In some cases, the “competition effect” dominates the “entry effect” and in other cases, it is vice versa. Hence, there is no monotone decreasing relationship between b and N .

6. **Lemma 3** $\frac{\partial b}{\partial N} < 0$.

Proof From the equilibrium bidding strategy, we have

$$\begin{aligned} b_N &= c + \frac{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} \int_c^{\bar{c}} (1-F(x))^{j-1} dx}{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(c))^{j-1}} \\ &= c + \int_c^{\bar{c}} \frac{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(x))^{j-1}}{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(c))^{j-1}} dx. \end{aligned}$$

Then if

$$\frac{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(x))^{j-1}}{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(c))^{j-1}}$$

decreases with N , then b_N decreases with N and the proof is done. Denote $1-F(x) = a$ and $1-F(c) = d$. Note that x takes values on (c, \bar{c}) , which implies $a < d$. Then,

$$\begin{aligned} b_N - b_{N-1} &= \frac{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(x))^{j-1}}{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(c))^{j-1}} \\ &\quad - \frac{\sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} (1-F(x))^{j-1}}{\sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} (1-F(c))^{j-1}} \\ &= \frac{\sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} d^{j-1} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} a^{j-1}}{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} d^{j-1} \sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} d^{j-1}} \\ &\quad - \frac{\sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} a^{j-1} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} d^{j-1}}{\sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} d^{j-1} \sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} d^{j-1}}. \end{aligned}$$

Therefore, the sign of $b_N - b_{N-1}$ is determined by

$$\begin{aligned} &\sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} d^{j-1} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} a^{j-1} \\ &- \sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} a^{j-1} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} d^{j-1}. \end{aligned} \quad (\text{A.6})$$

We now show that this term is greater than 0. The first term is

$$\begin{aligned}
& \sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} d^{j-1} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} a^{j-1} \\
= & \left[\binom{N}{1} q^* (1-q^*)^{N-1} d + \binom{N}{2} (q^*)^2 (1-q^*)^{N-2} d^2 + \dots + \binom{N}{N} (q^*)^N d^N \right] \\
& * \left[\binom{N-1}{1} q^* (1-q^*)^{N-2} a + \binom{N-1}{2} (q^*)^2 (1-q^*)^{N-3} a^2 + \dots + \binom{N-1}{N-1} (q^*)^{N-1} a^{N-1} \right] \\
= & \left[\binom{N}{1} \binom{N-1}{1} (q^*)^2 (1-q^*)^{2N-3} ad + \binom{N}{1} \binom{N-1}{2} (q^*)^3 (1-q^*)^{2N-4} a^2 d + \right. \\
& \quad \left. \dots + \binom{N}{1} \binom{N-1}{N-1} (q^*)^N (1-q^*)^{N-1} a^{N-1} d \right] \\
& + \left[\binom{N}{2} \binom{N-1}{1} (q^*)^3 (1-q^*)^{2N-4} ad^2 + \binom{N}{2} \binom{N-1}{2} (q^*)^4 (1-q^*)^{2N-5} a^2 d^2 + \right. \\
& \quad \left. \dots + \binom{N}{2} \binom{N-1}{N-1} (q^*)^{N+1} (1-q^*)^{N-2} a^{N-1} d^2 \right] \\
& + \dots \\
& + \left[\binom{N}{N} \binom{N-1}{1} (q^*)^{N+1} (1-q^*)^{N-2} ad^N + \binom{N}{N} \binom{N-1}{2} (q^*)^{N+2} (1-q^*)^{N-3} a^2 d^N + \right. \\
& \quad \left. \dots + \binom{N}{N} \binom{N-1}{N-1} (q^*)^{2N-1} a^{N-1} d^N \right]. \tag{A.7}
\end{aligned}$$

Note that after the expansion, equation (A.7) has $N(N-1)$ terms. It has N rows and each row has $N-1$ columns. Similarly, the second term can be expanded as

$$\begin{aligned}
& \sum_{j=2}^{N+1} \binom{N}{j-1} (q^*)^{j-1} (1-q^*)^{N+1-j} a^{j-1} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} d^{j-1} \\
= & \left[\binom{N}{1} q^* (1-q^*)^{N-1} a + \binom{N}{2} (q^*)^2 (1-q^*)^{N-2} a^2 + \dots + \binom{N}{N} (q^*)^N a^N \right] \\
& * \left[\binom{N-1}{1} q^* (1-q^*)^{N-2} d + \binom{N-1}{2} (q^*)^2 (1-q^*)^{N-3} d^2 + \dots + \binom{N-1}{N-1} (q^*)^{N-1} d^{N-1} \right] \\
= & \left[\binom{N}{1} \binom{N-1}{1} (q^*)^2 (1-q^*)^{2N-3} ad + \binom{N}{1} \binom{N-1}{2} (q^*)^3 (1-q^*)^{2N-4} ad^2 + \right. \\
& \quad \left. \dots + \binom{N}{1} \binom{N-1}{N-1} (q^*)^N (1-q^*)^{N-1} ad^{N-1} \right] \\
& + \left[\binom{N}{2} \binom{N-1}{1} (q^*)^3 (1-q^*)^{2N-4} a^2 d + \binom{N}{2} \binom{N-1}{2} (q^*)^4 (1-q^*)^{2N-5} a^2 d^2 + \right. \\
& \quad \left. \dots + \binom{N}{2} \binom{N-1}{N-1} (q^*)^{N+1} (1-q^*)^{N-2} a^2 d^{N-1} \right] \\
& + \dots \\
& + \left[\binom{N}{N} \binom{N-1}{1} (q^*)^{N+1} (1-q^*)^{N-2} a^N d + \binom{N}{N} \binom{N-1}{2} (q^*)^{N+2} (1-q^*)^{N-3} a^N d^2 + \right. \\
& \quad \left. \dots + \binom{N}{N} \binom{N-1}{N-1} (q^*)^{2N-1} a^N d^{N-1} \right]. \tag{A.8}
\end{aligned}$$

Again after expansion, equation (A.8) has $N(N-1)$ terms. It has N rows and each row has $N-1$ columns. Now, we have two expressions (A.7) and (A.8), each of which has $N(N-1)$

terms. Those terms are best indexed by the powers of a and d . In the first expression, we have terms with powers like

$$\begin{aligned} & ad, ad^2, \dots, ad^N; \\ & a^2d, a^2d^2, \dots, a^2d^N; \\ & \dots \\ & a^{N-1}d, a^{N-1}d^2, \dots, a^{N-1}d^N, \end{aligned}$$

while in the second expression, we have terms like

$$\begin{aligned} & ad, ad^2, \dots, ad^{N-1}; \\ & a^2d, a^2d^2, \dots, a^2d^{N-1}; \\ & \dots \\ & a^N d, a^N d^2, \dots, a^N d^{N-1}. \end{aligned}$$

Furthermore, notice that a term with power $a^i d^j$ in expression (A.7) has the coefficient as

$$\binom{N}{j} \binom{N-1}{i} (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)},$$

while a term has power $a^i d^j$ in expression (A.8) has the coefficient as

$$\binom{N}{i} \binom{N-1}{j} (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)}.$$

Therefore, when we subtract the second expression from the first expression in equation (A.6), those terms with power indices $i = j$ cancel with each other. For a generic i and j with $i \leq N-1, j \leq N-1$ and without loss of generality, $i < j$, we can combine the four terms

$$\begin{aligned} & \binom{N}{j} \binom{N-1}{i} (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} a^i d^j + \binom{N}{i} \binom{N-1}{j} (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} a^j d^i \\ & - \binom{N}{i} \binom{N-1}{j} (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} a^i d^j - \binom{N}{j} \binom{N-1}{i} (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} a^j d^i \\ = & \left[\binom{N}{j} \binom{N-1}{i} - \binom{N}{i} \binom{N-1}{j} \right] (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} a^i d^j \\ & + \left[\binom{N}{i} \binom{N-1}{j} - \binom{N}{j} \binom{N-1}{i} \right] (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} a^j d^i \\ = & \left[\binom{N}{j} \binom{N-1}{i} - \binom{N}{i} \binom{N-1}{j} \right] (q^*)^{i+j} (1-q^*)^{2N-1-(i+j)} [a^i d^j - a^j d^i]. \end{aligned}$$

Since $i < j$,

$$a^i d^j - a^j d^i = a^i d^i [d^{j-i} - a^{j-i}] > 0 \text{ since } a < d.$$

Also,

$$\begin{aligned} & \binom{N}{j} \binom{N-1}{i} - \binom{N}{i} \binom{N-1}{j} \\ = & \frac{N!}{(N-j)!j!} \frac{(N-1)!}{(N-1-i)!i!} - \frac{N!}{(N-i)!i!} \frac{(N-1)!}{(N-1-j)!j!} \\ = & \frac{N!(N-1)!}{j!i!} \left[\frac{1}{(N-j)!(N-1-i)!} - \frac{1}{(N-i)!(N-1-j)!} \right] \\ > & 0 \text{ since } i < j. \end{aligned}$$

Therefore, the combination of the above four terms will yield a value greater than 0. Now, the rest are the terms with powers

$$ad^N; a^2 d^N \dots a^{N-1} d^N.$$

in expression (A.7) and terms with powers

$$a^N d, a^N d^2, \dots a^N d^{N-1}.$$

in expression (A.8). Then each pair, for example,

$$\begin{aligned} & \binom{N}{N} \binom{N-1}{1} (q^*)^{N+1} (1-q^*)^{N-2} ad^N - \binom{N}{N} \binom{N-1}{1} (q^*)^{N+1} (1-q^*)^{N-2} a^N d \\ = & \binom{N}{N} \binom{N-1}{1} (q^*)^{N+1} (1-q^*)^{N-2} ad(d^{N-1} - a^{N-1}) > 0 \text{ since } a < d. \end{aligned}$$

Thus, we make a combination of the $2N(N-1)$ terms in such a way that each resulting term is positive. This completes the proof.

7. **Lemma 4** $\frac{\partial b}{\partial q^*} < 0$.

Proof From the equilibrium bidding strategy, we have

$$b_N = c + \frac{\sum_{j=2}^N P_B(n=j|N \geq n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx}{\sum_{j=2}^N P_B(n=j|N \geq n \geq 2) (1-F(c))^{j-1}}.$$

Then

$$\begin{aligned} \frac{\partial b}{\partial q^*} \propto & \sum_{j=2}^N \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} \int_c^{\bar{c}} (1-F(x))^{j-1} dx \sum_{j=2}^N P_B(n=j|N \geq n \geq 2) (1-F(c))^{j-1} \\ & - \sum_{j=2}^N P_B(n=j|N \geq n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx \sum_{j=2}^N \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} (1-F(c))^{j-1}. \end{aligned}$$

After expansion, $\frac{\partial b}{\partial q^*}$ becomes a sum of $(N-1)(N-2)/2$ terms. Each term can be expressed by

$$\begin{aligned}
& \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} \int_c^{\bar{c}} (1-F(x))^{j-1} dx P_B(n=k|N \geq n \geq 2) (1-F(c))^{k-1} \\
& - P_B(n=k|N \geq n \geq 2) \int_c^{\bar{c}} (1-F(x))^{k-1} dx \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} (1-F(c))^{j-1} \\
& + \frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} \int_c^{\bar{c}} (1-F(x))^{k-1} dx P_B(n=j|N \geq n \geq 2) (1-F(c))^{j-1} \\
& - P_B(n=j|N \geq n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx \frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} (1-F(c))^{k-1}
\end{aligned} \tag{A.9}$$

for $2 \leq j < k \leq N$. Therefore, if each term of (A.9) is negative, then the proof $\frac{\partial b}{\partial q^*} < 0$ is done. Simplifying it, we get

$$\begin{aligned}
& \int_c^{\bar{c}} (1-F(x))^{j-1} dx (1-F(c))^{k-1} \\
& * \left[\frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} P_B(n=k|N \geq n \geq 2) - \frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} P_B(n=j|N \geq n \geq 2) \right] \\
& + \int_c^{\bar{c}} (1-F(x))^{k-1} dx (1-F(c))^{j-1} \\
& * \left[\frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} P_B(n=j|N \geq n \geq 2) - \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} P_B(n=k|N \geq n \geq 2) \right] \\
& = \left[\frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} P_B(n=k|N \geq n \geq 2) - \frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} P_B(n=j|N \geq n \geq 2) \right] \\
& * \left[\int_c^{\bar{c}} (1-F(x))^{j-1} dx (1-F(c))^{k-1} - \int_c^{\bar{c}} (1-F(x))^{k-1} dx (1-F(c))^{j-1} \right].
\end{aligned} \tag{A.10}$$

Here,

$$\begin{aligned}
& \left[\int_c^{\bar{c}} (1-F(x))^{j-1} dx (1-F(c))^{k-1} - \int_c^{\bar{c}} (1-F(x))^{k-1} dx (1-F(c))^{j-1} \right] \\
& = \left[\int_c^{\bar{c}} (1-F(x))^{j-1} (1-F(c))^{k-1} dx - \int_c^{\bar{c}} (1-F(x))^{k-1} (1-F(c))^{j-1} dx \right] \\
& = \int_c^{\bar{c}} (1-F(x))^{j-1} (1-F(c))^{j-1} [(1-F(c))^{k-j} - (1-F(x))^{k-j}] dx > 0
\end{aligned}$$

Also, since

$$\begin{aligned}
P_B(n=k|N \geq n \geq 2) &= \binom{N-1}{k-1} (q^*)^{k-1} (1-q^*)^{N-k} / \left(1 - (1-q^*)^N \right) \\
&= \frac{(N-1)!}{(N-k)!(k-1)!} \frac{(q^*)^{k-1} (1-q^*)^{N-k}}{1-(1-q^*)^N} \\
&= \frac{(N-1)!(N-j)(N-j-1)\dots(N-k)}{(N-j)!(j-1)!(k-1)(k-2)\dots(j)} \frac{(q^*)^{j-1} (1-q^*)^{N-j} (q^*)^{k-j} (1-q^*)^{j-k}}{1-(1-q^*)^N} \\
&= \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} P_B(n=j|N \geq n \geq 2) (q^*)^{k-j} (1-q^*)^{j-k}.
\end{aligned}$$

we have,

$$\begin{aligned}
& \frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} \\
& = \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} (q^*)^{k-j} (1-q^*)^{j-k} \\
& + \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} P_B(n=j|N \geq n \geq 2) \\
& * \left[(k-j)(q^*)^{k-j-1} (1-q^*)^{j-k} + (k-j)(q^*)^{k-j} (1-q^*)^{j-k-1} \right].
\end{aligned}$$

Thus,

$$\begin{aligned}
& \left[\frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} P_B(n=k|N \geq n \geq 2) - \frac{\partial P_B(n=k|N \geq n \geq 2)}{\partial q^*} P_B(n=j|N \geq n \geq 2) \right] \\
&= \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} P_B(n=j|N \geq n \geq 2) \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} (q^*)^{k-j} (1-q^*)^{j-k} \\
&\quad - P_B(n=j|N \geq n \geq 2) \frac{\partial P_B(n=j|N \geq n \geq 2)}{\partial q^*} \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} (q^*)^{k-j} (1-q^*)^{j-k} \\
&\quad - P_B(n=j|N \geq n \geq 2) \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} P_B(n=j|N \geq n \geq 2) \\
&\quad * [(k-j)(q^*)^{k-j-1}(1-q^*)^{j-k} + (k-j)(q^*)^{k-j}(1-q^*)^{j-k-1}] \\
&= - [P_B(n=j|N \geq n \geq 2)]^2 \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} (k-j)(q^*)^{k-j-1}(1-q^*)^{j-k-1} [(1-q^*) + q^*] \\
&= - [P_B(n=j|N \geq n \geq 2)]^2 \frac{(N-j)(N-j-1)\dots(N-k)}{(k-1)(k-2)\dots(j)} (k-j)(q^*)^{k-j-1}(1-q^*)^{j-k-1} < 0.
\end{aligned}$$

Therefore, (A.10) is negative and this completes the proof.

8. **Proof of Proposition 4** As proved by lemma 5 below, the expected winning bid (procurement cost) from the government's point of view can be written as

$$\begin{aligned}
W(q^*, N) &= \sum_{v=2}^N \binom{N}{v} \frac{(q^*)^v (1-q^*)^{N-v}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} E \left[v(1-F(c))^{v-1}c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] \\
&= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) E \left[v(1-F(c))^{v-1}c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] \\
&= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) \int_{\underline{c}}^{\bar{c}} \left[v(1-F(c))^{v-1}c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] f(c) dc
\end{aligned} \tag{A.11}$$

where $P_G(n=v|n \geq 2, q^*) = \binom{N}{v} \frac{(q^*)^v (1-q^*)^{N-v}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}}$ is the probability that there will be j actual bidders from the government's point of view.

N affects W in two channels:

$$\frac{dW}{dN} = \frac{\partial W}{\partial N} + \frac{\partial W}{\partial q^*} \frac{\partial q^*}{\partial N}.$$

We first examine the entry effect, that is, $\frac{\partial W}{\partial q^*} \frac{\partial q^*}{\partial N}$. Since $\frac{\partial q^*}{\partial N} < 0$ is proved by Proposition 2, if $\frac{\partial W}{\partial q^*} < 0$, then the entry effect is positive as desired.

Denote the term in equation (A.11) $\int_{\underline{c}}^{\bar{c}} \left[v(1-F(c))^{v-1}c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] f(c) dc = E_v$.

Note that

$$\begin{aligned}
& \frac{\partial P_G(n=v|n \geq 2, q^*)}{\partial q^*} \propto [vq^{*v-1}(1-q^*)^{N-v} - (N-v)q^{*v}(1-q^*)^{N-v-1}] \\
& * [1 - (1-q^*)^N - Nq^*(1-q^*)^{N-1}] - N(N-1)q^{*v+1}(1-q^*)^{2N-v-2} \\
& \propto [v(1-q^*) - (N-v)q^*] * [1 - (1-q^*)^N - Nq^*(1-q^*)^{N-1}] \\
& \quad - N(N-1)q^{*2}(1-q^*)^{N-1} \\
& = (v - Nq^*) * [1 - (1-q^*)^N - Nq^*(1-q^*)^{N-1}] - N(N-1)q^{*2}(1-q^*)^{N-1}.
\end{aligned}$$

Therefore, $\frac{\partial P_G(n=v|n \geq 2, q^*)}{\partial q^*} < 0$ if $v < \frac{N(N-1)q^{*2}(1-q^*)^{N-1}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} + Nq^*$ and $\frac{\partial P_G(n=v|n \geq 2, q^*)}{\partial q^*} \geq 0$ otherwise. Thus, without loss of generality, there exists an integer $N > Q > 2$ such that for $2 \leq v \leq Q$, $P_G(n = v|n \geq 2, q^*)$ decreases in q^* and for $v > Q$, $P_G(n = v|n \geq 2, q^*)$ increases in q^* .

Now, suppose q^* increases from q_1^* to q_2^* with $0 < q_1^* < q_2^* < 1$, the changes in $W(q^*, N)$ can be written as

$$\begin{aligned}
W(q_2^*, N) - W(q_1^*, N) &= \sum_{v=2}^N P_G(n = v|n \geq 2, q_2^*) E_v - \sum_{v=2}^N P_G(n = v|n \geq 2, q_1^*) E_v \\
&= [P_G(n = 2|n \geq 2, q_2^*) - P_G(n = 2|n \geq 2, q_1^*)] E_2 \\
&\quad + \sum_{v=3}^Q [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] E_v \\
&\quad + [P_G(n = Q + 1|n \geq 2, q_2^*) - P_G(n = Q + 1|n \geq 2, q_1^*)] E_{Q+1} \\
&\quad + \sum_{v=Q+2}^N [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] E_v \\
&\leq [P_G(n = 2|n \geq 2, q_2^*) - P_G(n = 2|n \geq 2, q_1^*)] E_2 \\
&\quad + \sum_{v=3}^Q [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] E_{Q+1} \\
&\quad + [P_G(n = Q + 1|n \geq 2, q_2^*) - P_G(n = Q + 1|n \geq 2, q_1^*)] E_{Q+1} \\
&\quad + \sum_{v=Q+2}^N [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] E_{Q+1} \\
&= [P_G(n = 2|n \geq 2, q_2^*) - P_G(n = 2|n \geq 2, q_1^*)] E_2 \\
&\quad + \sum_{v=3}^N [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] E_{Q+1} \\
&\leq [P_G(n = 2|n \geq 2, q_2^*) - P_G(n = 2|n \geq 2, q_1^*)] E_2 \\
&\quad + \sum_{v=3}^N [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] E_2 \\
&= E_2 \sum_{v=2}^N [P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] = 0.
\end{aligned}$$

The first inequality follows from the facts that E_v is decreasing in v as proved in Lemma 6 below and $P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*) < 0$ for $v \leq Q$ and $[P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*)] \geq 0$ for $v > Q$. The second inequality follows from the additional fact that $\sum_{v=3}^N P_G(n = v|n \geq 2, q_2^*) - P_G(n = v|n \geq 2, q_1^*) > 0$. This is because $\sum_{v=2}^N P_G(n = v|n \geq 2, q_2^*) = \sum_{v=2}^N P_G(n = v|n \geq 2, q_1^*) = 1$ and $P_G(n = 2|n \geq 2, q_2^*) - P_G(n = 2|n \geq 2, q_1^*) < 0$. This proves that $\frac{\partial W}{\partial q^*} < 0$ and hence the entry effect $\frac{\partial W}{\partial q^*} \frac{\partial q^*}{\partial N} > 0$. This completes proof of (ii).

We now turn to the competition effect, that is, $\frac{\partial W}{\partial N}$. Since

$$\begin{aligned}
1 &= \sum_{v=2}^N P_G(n = v|N \geq n \geq 2, q^*) = \sum_{v=2}^N \binom{N}{v} \frac{(q^*)^v (1-q^*)^{N-v}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \\
&= \sum_{v=2}^{N+1} \binom{N+1}{v} \frac{(q^*)^v (1-q^*)^{N+1-v}}{1-(1-q^*)^{N+1} - (N+1)q^*(1-q^*)^N} = \sum_{v=2}^{N+1} P_G(n = v|N + 1 \geq n \geq 2, q^*) ,
\end{aligned}$$

therefore

$$\begin{aligned} \frac{q^{*N+1}}{1-(1-q^*)^{N+1}-(N+1)q^*(1-q^*)^N} &= \sum_{v=2}^N \left[\frac{\binom{N}{v}(q^*)^v(1-q^*)^{N-v}}{1-(1-q^*)^N-Nq^*(1-q^*)^{N-1}} - \frac{\binom{N+1}{v}(q^*)^v(1-q^*)^{N+1-v}}{1-(1-q^*)^{N+1}-(N+1)q^*(1-q^*)^N} \right] \\ &= \sum_{v=2}^N B_v \end{aligned} \quad (\text{A.12})$$

where $B_v = \frac{\binom{N}{v}(q^*)^v(1-q^*)^{N-v}}{1-(1-q^*)^N-Nq^*(1-q^*)^{N-1}} - \frac{\binom{N+1}{v}(q^*)^v(1-q^*)^{N+1-v}}{1-(1-q^*)^{N+1}-(N+1)q^*(1-q^*)^N}$. Then,

$$\begin{aligned} W_N - W_{N+1} &= \sum_{v=2}^N P_G(n=v|N \geq n \geq 2, q^*)E_v - \sum_{v=2}^{N+1} P_G(n=v|N+1 \geq n \geq 2, q^*)E_v \\ &= \sum_{v=2}^N [P_G(n=v|N \geq n \geq 2, q^*) - P_G(n=v|N+1 \geq n \geq 2, q^*)] E_v \\ &\quad - \frac{q^{*N+1}}{1-(1-q^*)^{N+1}-(N+1)q^*(1-q^*)^N} E_{N+1} \\ &= \sum_{v=2}^N B_v E_v - \sum_{v=2}^N B_v E_{N+1} = \sum_{v=2}^N B_v (E_v - E_{N+1}) \end{aligned}$$

If for all $v = 2, \dots, N$, $B_v > 0$, then $W_N - W_{N+1} \geq 0$ because $E_v - E_{N+1} \geq 0$ from Lemma 6.

Otherwise, note that

$$B_v = \binom{N}{v} (q^*)^v (1-q^*)^{N-v} \left[\frac{1}{1-(1-q^*)^N-Nq^*(1-q^*)^{N-1}} - \frac{\frac{N+1}{N-v+1}}{\frac{1}{1-q^*}-(1-q^*)^N-(N+1)q^*(1-q^*)^{N-1}} \right].$$

As v increases, $\frac{\frac{N+1}{N-v+1}}{\frac{1}{1-q^*}-(1-q^*)^N-(N+1)q^*(1-q^*)^{N-1}}$ increases. Thus there exists a v_0 such that for $v < v_0$, $B_v > 0$ and for $v \geq v_0$, $B_v < 0$. Thus,

$$\begin{aligned} W_N - W_{N+1} &\geq \sum_{v=2}^{v_0-1} B_v (E_{v_0-1} - E_{N+1}) + \sum_{v_0}^N B_v (E_{v_0-1} - E_{N+1}) \\ &\geq \sum_{v=2}^N B_v (E_{v_0-1} - E_{N+1}) = \frac{q^{*N+1}}{1-(1-q^*)^{N+1}-(N+1)q^*(1-q^*)^N} (E_{v_0-1} - E_{N+1}) \geq 0 \end{aligned}$$

where we use $\sum_{v=2}^N B_v = \frac{q^{*N+1}}{1-(1-q^*)^{N+1}-(N+1)q^*(1-q^*)^N}$ from (A.12).

9. **Lemma 5** $W(q^*, N) = \sum_{v=2}^N \binom{N}{v} \frac{(q^*)^v(1-q^*)^{N-v}}{1-(1-q^*)^N-Nq^*(1-q^*)^{N-1}} E \left[\begin{array}{l} v(1-F(c))^{v-1}c \\ +v \int_c^{\bar{c}} (1-F(x))^{v-1} dx \end{array} \right].$

Proof

$$\begin{aligned} W(q^*, N) &= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) \int_c^{\bar{c}} v [1-F(c)]^{v-1} f(c) s(c|N, q^*) dc \\ &= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) \int_c^{\bar{c}} v [1-F(c)]^{v-1} f(c) * \left[c + \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx}{\sum_{j=2}^N P_B(n=j|n \geq 2) (1-F(c))^{j-1}} \right] dc \\ &= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) \int_c^{\bar{c}} v [1-F(c)]^{v-1} f(c) c dc \\ &\quad + \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) \int_c^{\bar{c}} v [1-F(c)]^{v-1} f(c) \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \int_c^{\bar{c}} (1-F(x))^{j-1} dx}{\sum_{j=2}^N P_B(n=j|n \geq 2) (1-F(c))^{j-1}} dc \end{aligned}$$

$$\begin{aligned}
&= \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) v [1 - F(c)]^{v-1} f(c) dc \\
&+ \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) v [1 - F(c)]^{v-1} f(c) \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx}{\sum_{j=2}^N P_B(n=j|n \geq 2) (1-F(c))^{j-1}} dc \\
&= \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) v [1 - F(c)]^{v-1} f(c) dc \\
&+ \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N \frac{\binom{N}{v} (q^*)^v (1-q^*)^{N-v}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} v [1 - F(c)]^{v-1} f(c) \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx}{\sum_{j=2}^N P_B(n=j|n \geq 2) (1-F(c))^{j-1}} dc \\
&= \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N \frac{\binom{N}{v} (q^*)^v (1-q^*)^{N-v}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} v [1 - F(c)]^{v-1} f(c) dc \\
&+ \frac{Nq^*}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N \binom{N-1}{v-1} (q^*)^{v-1} (1-q^*)^{N-v} [1 - F(c)]^{v-1} f(c) \\
&* \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx}{\frac{1}{1-(1-q^*)^{N-1}} \sum_{j=2}^N \binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j} (1-F(c))^{j-1}} dc \\
&= \frac{Nq^*}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N \binom{N-1}{v-1} (q^*)^{v-1} (1-q^*)^{N-v} [1 - F(c)]^{v-1} f(c) dc \\
&+ \frac{Nq^*(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \int_{\underline{c}}^{\bar{c}} f(c) \sum_{j=2}^N P_B(n=j|n \geq 2) \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx dc \\
&= \frac{Nq^*(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \int_{\underline{c}}^{\bar{c}} \sum_{v=2}^N \frac{\binom{N-1}{v-1} (q^*)^{v-1} (1-q^*)^{N-v}}{(1-(1-q^*)^{N-1})} [1 - F(c)]^{v-1} f(c) dc \\
&+ \frac{Nq^*(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \int_{\underline{c}}^{\bar{c}} f(c) \sum_{j=2}^N P_B(n=j|n \geq 2) \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx dc \\
&= \frac{Nq^*(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \left[\int_{\underline{c}}^{\bar{c}} \sum_{j=2}^N P_B(n=j|n \geq 2) (1-F(c))^{j-1} f(c) dc \right. \\
&\quad \left. + \int_{\underline{c}}^{\bar{c}} f(c) \sum_{j=2}^N P_B(n=j|n \geq 2) \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx dc \right] \\
&= \frac{Nq^*(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \sum_{j=2}^N P_B(n=j|n \geq 2) \left[\int_{\underline{c}}^{\bar{c}} (1-F(c))^{j-1} f(c) dc \right. \\
&\quad \left. + \int_{\underline{c}}^{\bar{c}} \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx f(c) dc \right] \\
&= \frac{Nq^*(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \sum_{j=2}^N P_B(n=j|n \geq 2) E \left[(1-F(c))^{j-1} c + \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx \right] \\
&= \frac{N(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \sum_{j=2}^N q^* \frac{\binom{N-1}{j-1} (q^*)^{j-1} (1-q^*)^{N-j}}{1-(1-q^*)^{N-1}} E \left[(1-F(c))^{j-1} c + \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx \right] \\
&= \frac{N(1-(1-q^*)^{N-1})}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} \sum_{j=2}^N \binom{N}{j} \frac{(q^*)^j (1-q^*)^{N-j}}{1-(1-q^*)^{N-1}} \frac{j}{N} E \left[(1-F(c))^{j-1} c + \int_{\underline{c}}^{\bar{c}} (1-F(x))^{j-1} dx \right] \\
&= \sum_{v=2}^N \binom{N}{v} \frac{(q^*)^v (1-q^*)^{N-v}}{1-(1-q^*)^N - Nq^*(1-q^*)^{N-1}} E \left[v(1-F(c))^{v-1} c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] \\
&= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) E \left[v(1-F(c))^{v-1} c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] \\
&= \sum_{v=2}^N P_G(n=v|n \geq 2, q^*) \int_{\underline{c}}^{\bar{c}} \left[v(1-F(c))^{v-1} c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] f(c) dc.
\end{aligned}$$

10. **Lemma 6** Let $E_v = \int_{\underline{c}}^{\bar{c}} \left[v(1-F(c))^{v-1} c + v \int_{\underline{c}}^{\bar{c}} (1-F(x))^{v-1} dx \right] f(c) dc$. Then E_v is decreasing in v if $c + F(c)/f(c)$ is increasing in c .

Proof Integration by parts gives $E_v = \int_c^{\bar{c}} v(1 - F(c))^{v-1} f(c)[c + F(c)/f(c)]dc$. Let $G_v(c) \equiv 1 - (1 - F(c))^v$. Because $G_v(c) < G_{v+1}(c)$, $G_v(c)$ (first order) stochastically dominates $G_{v+1}(c)$. Noting that $E_v = E_{G_v}[c + F(c)/f(c)]$, and invoking a standard property of stochastic dominance (see e.g., Appendix B in Krishna (2002)), we have that if $c + F(c)/f(c)$ is increasing in c , $E_v \geq E_{v+1}$, that is E_v is decreasing in v .

11. Derivation of (11)

$$\begin{aligned} & \frac{\partial s(c) \sum_{j=2}^N P_B(n = j | n \geq 2) [1 - F(c)]^{j-1}}{\partial c} \\ &= -c \sum_{j=2}^N P_B(n = j | n \geq 2) (j-1) [1 - F(c)]^{j-2} f(c) \end{aligned}$$

Therefore, integration on both sides:

$$\begin{aligned} & s(c) \sum_{j=2}^N P_B(n = j | n \geq 2) [1 - F(c)]^{j-1} \Big|_c^{\bar{c}} \\ &= - \int_c^{\bar{c}} x \sum_{j=2}^N P_B(n = j | n \geq 2) (j-1) [1 - F(x)]^{j-2} f(x) dx + \text{constant} \quad (28) \end{aligned}$$

Assuming that $f(c|\mu_1) = \frac{1}{\exp(\mu_1)} \exp(-\frac{1}{\exp(\mu_1)}c)$ on $(0, \infty)$ and therefore $1 - F(c) = \exp(-\frac{1}{\exp(\mu_1)}c)$, equation (A.9) becomes

$$\begin{aligned} & -s(c) \sum_{j=2}^N P_B(n = j | n \geq 2) [1 - F(c)]^{j-1} \\ &= - \sum_{j=2}^N P_B(n = j | n \geq 2) (j-1) * \int_c^{\infty} x \exp(-\frac{j-2}{\exp(\mu_1)}x) \frac{1}{\exp(\mu_1)} \exp(-\frac{1}{\exp(\mu_1)}x) dx + \text{constant} \\ &= \sum_{j=2}^N P_B(n = j | n \geq 2) \int_c^{\infty} x d \exp(-\frac{j-1}{\exp(\mu_1)}x) + \text{constant} \\ &= - \sum_{j=2}^N P_B(n = j | n \geq 2) c \exp(-\frac{j-1}{\exp(\mu_1)}c) \\ &= - \sum_{j=2}^N \frac{\exp(\mu_1)}{j-1} P_B(n = j | n \geq 2) \exp(-\frac{j-1}{\exp(\mu_1)}c) + \text{constant} \end{aligned}$$

Therefore,

$$\begin{aligned} s(c) &= \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) c \exp(-\frac{j-1}{\exp(\mu_1)}c)}{\sum_{j=2}^N P_B(n=j|n \geq 2) [1-F(c)]^{j-1}} + \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \exp(-\frac{j-1}{\exp(\mu_1)}c) \frac{\exp(\mu_1)}{j-1}}{\sum_{j=2}^N P_B(n=j|n \geq 2) [1-F(c)]^{j-1}} \\ &+ \frac{\text{constant}}{\sum_{j=2}^N P_B(n=j|n \geq 2) [1-F(c)]^{j-1}} \\ &= c + \frac{\sum_{j=2}^N P_B(n=j|n \geq 2) \exp(-\frac{j-1}{\exp(\mu_1)}c) \frac{\exp(\mu_1)}{j-1}}{\sum_{j=2}^N P_B(n=j|n \geq 2) [1-F(c)]^{j-1}} \end{aligned}$$

with the boundary condition that $\lim_{c \rightarrow \infty} (s(c) - c) = 0$.

12. **Derivation of (13)** The ex ante expected payoff for the contractor in the mixed-strategy model is:

$$\begin{aligned}
\int_0^\infty E\pi(b, c)f(c)dc &= \\
\int_0^\infty \sum_{j=2}^N P_B(n = j|n \geq 2) \exp\left(-\frac{j-1}{\exp(\mu_1)}c\right) \frac{\exp(\mu_1)}{j-1} \frac{1}{\exp(\mu_1)} \exp\left(-\frac{1}{\exp(\mu_1)}c\right)dc &= \\
= \sum_{j=2}^N P_B(n = j|n \geq 2) \frac{1}{j-1} \int_0^\infty \exp\left(-\frac{j}{\exp(\mu_1)}c\right)dc &= \\
= \sum_{j=2}^N P_B(n = j|n \geq 2) \frac{1}{j-1} \frac{\exp(\mu_1)}{j} &=
\end{aligned}$$

Appendix II: Details of Estimation Method

A. Bayesian Nonparametric Density Estimation

Our objective here for the nonparametric Bayesian estimation is to update the number of components in approximating mixture of normals in (20) and the mean and variance for each component given a set of observations $e_{new,\ell}$. The normal kernel is chosen here to follow Escobar and West (1995), Hirano (2002) and Hasegawa and Kozumi (2003). The density for $e_{new,\ell}$ conditional on the parameters in the normal density can be written as

$$q(e_{new,\ell}|\theta_\ell) = q(e_{new,\ell}|d_\ell, \sigma_\ell^2) = \phi(e_{new,\ell}|d_\ell, \sigma_\ell^2). \quad (\text{A.12})$$

We further assume that $\theta_\ell = (d_\ell, \sigma_\ell^2)$, $\ell = 1, 2, \dots, L$ is a sample from some unknown distribution P on the space (Ω, Γ) and Γ is a σ -field of subsets of Ω . Hence, conditional on P , the density for $e_{new,\ell}$ is

$$s(e_{new,\ell}|P) = \int q(e_{new,\ell}|\theta_\ell)dP. \quad (\text{A.13})$$

This gives us a nonparametric class of distributions. How rich this class is depends on the choice of q and P . For Bayesian analysis, we need to specify a prior for P , which in turn implies a prior for $s(e_{new,\ell}|P)$. If the prior for P has unit mass at some point θ^* , then the prior for $s(e_{new,\ell}|P)$ is just equal to $q(e_{new,\ell}|\theta^*)$, which is the same as a parametric specification of s . In our case, it is a normal density with parameters $\theta^* = (d^*, \sigma^{2*})$. If P is assumed to have a Dirichlet process prior as introduced by Ferguson (1973, 1974), then the resulting prior for $s(e_{new,\ell}|P)$ becomes a prior structure that can be used for any random smooth density and hence can be regarded as a nonparametric specification of s .

To further understand why the use of the infinite countable normal mixture model (20) and the Dirichlet process prior for P can be regarded as a nonparametric specification of the density of $e_{new,\ell}$, we need to know what a Dirichlet process is and what is the link between a Dirichlet process prior for P and the infinite countable normal mixture model. The Dirichlet process is a probability measure on the space of all distributions and can be used as a prior distribution on the space of probability distributions. Let τP_0 be a finite non-null measure on (Ω, Γ) where P_0 is a proper base probability distribution and τ is a precision parameter. Then, a stochastic process P is a Dirichlet process if, for any given partition, A_1, \dots, A_q of the parameter space Ω , the random vector

$(P(A_1), \dots, P(A_q))$ has a Dirichlet distribution with a parameter vector $(\tau P_0(A_1), \dots, \tau P_0(A_q))$. We denote $\text{DP}(\tau P_0)$ for a Dirichlet process with base measure τP_0 for the rest of the paper.

From the previous paragraph, it is clear that the Dirichlet process is a probability distribution on the space of probability distributions on (Ω, Γ) and selects a discrete distribution with probability one. The discreteness of the probability distribution selected by the Dirichlet process seems to be unsuitable for modelling smooth densities such as that of $e_{new, \ell}$. However, it can be related to the infinite countable normal mixture model as noted by Ferguson (1983), Lo (1984) and further explained by Ghosal, Ghosh and Ramamoorthi (1999). This can be seen most clearly in the representation derived by Sethuraman (1994). Let w_1, w_2, \dots be i.i.d. $\text{Beta}(1, \tau)$ and let $p_1 = w_1, p_j = w_j \prod_{i=1}^{j-1} (1 - w_i), j > 1$. Finally, let $\theta_1, \theta_2, \dots$ be i.i.d. from P_0 . Then if $P(d\theta) = \sum_{j=1}^{\infty} p_j 1_{\theta_j}(d\theta)$ where $1_{\theta_j}(A) = 1$ if $\theta_j \in A$ and $1_{\theta_j}(A) = 0$ otherwise. Then, it can be easily shown that $P \sim \text{DP}(\tau P_0)$. With (A.13), this implies

$$s(e_{new, \ell} | P) = \sum_{j=1}^{\infty} p_j q(e_{new, \ell} | \theta_j) = \sum_{j=1}^{\infty} p_j \phi(e_{new, \ell} | d_j, \sigma_j^2) = g(\cdot) \quad (\text{A.14})$$

This demonstrates that the infinite countable mixture of normals is just an alternative representation of the density for $e_{new, \ell}$ under the Dirichlet process prior for P and a normal kernel.

Furthermore, Escobar (1994) shows that with the θ_ℓ 's as the latent variables, integrating P over its prior distribution $\text{DP}(\tau P_0)$ gives the sequence of conditional density for $\theta_\ell, \ell = 1, \dots, L$ as the following

$$\theta_\ell | \theta_1, \theta_2, \dots, \theta_{\ell-1} = \begin{cases} = \theta_j & \text{with probability } \frac{1}{\tau + \ell - 1} \\ \sim P_0 & \text{with probability } \frac{\tau}{\tau + \ell - 1} \end{cases} \quad (\text{A.15})$$

where $\theta_1 \sim P_0$. As a result, θ_ℓ 's or $e_{new, \ell}$'s are partitioned into m_c groups such that all $e_{new, \ell}$ in the same group have the same θ_ℓ while those in different groups differ. These m_c distinct values θ_ℓ are a sample from P_0 . The degree of clustering is determined by τ . Small values of τ will tend to generate a sample with all the data points located in a few clusters, whereas larger values of τ will give a more diverse sample.

Also, using the Dirichlet process prior leads to another useful set of conditional distributions for θ_ℓ . Let $\xi_j, j = 1, 2, \dots, m_c$ denote the m_c distinct values of θ_ℓ and

$$\theta^{(\ell)} = (\theta_1, \dots, \theta_{\ell-1}, \theta_{\ell+1}, \dots, \theta_L) \quad (\text{A.16})$$

be the set of values of θ for units other than ℓ . Furthermore, let superscript (ℓ) refer to variables defined on all units other than ℓ and hence $\xi^{(\ell)} = (\xi_1^{(\ell)}, \dots, \xi_{m_c^{(\ell)}}^{(\ell)})$ are the distinct values among $(\theta_1, \dots, \theta_{\ell-1}, \theta_{\ell+1}, \dots, \theta_L)$. Escobar (1994) shows that using the Dirichlet process prior leads to the following conditional distribution for P ,

$$P|\theta^{(\ell)}, P_0 \sim DP(\tau P_0 + \sum_{j=1}^{m_c^{(\ell)}} n_j^{(\ell)} 1(\xi_j^{(\ell)})) \quad (\text{A.17})$$

where $n_j^{(\ell)}$ is the number of θ_ℓ taking the value of $\xi_j^{(\ell)}$ and $1(\xi_j^{(\ell)})$ represents unit point mass at $\theta_\ell = \xi_j^{(\ell)}$. This in turn implies that

$$\theta_\ell|\theta^{(\ell)}, P_0 \sim E(P|\theta^{(\ell)}, P_0) \sim \frac{\tau}{\tau + L - 1} P_0 + \frac{1}{\tau + L - 1} \sum_{j=1}^{m_c^{(\ell)}} n_j^{(\ell)} 1(\xi_j^{(\ell)}). \quad (\text{A.18})$$

is the prior distribution of θ_ℓ conditional on $\theta^{(\ell)}$ and P_0 .

To complete the model specification, we choose the base prior distribution P_0 to be the conjugate normal/inverse-gamma distribution

$$dP_0(d_\ell, \sigma_\ell^2) \propto \text{Normal}(\mu_0, \tau_0 \sigma_\ell^2) IG\left(\frac{N_0}{2}, \frac{R_0}{2}\right) \quad (\text{A.19})$$

where $IG\left(\frac{N_0}{2}, \frac{R_0}{2}\right)$ denotes the inverse-gamma distribution with parameters $\frac{N_0}{2}$ and $\frac{R_0}{2}$. Finally, as in Escobar and West (1995), τ is assumed to follow a gamma distribution

$$\tau \sim \text{gamma}(d_1, d_2). \quad (\text{A.20})$$

One way of sampling θ_ℓ from its posterior distribution and determines the number of components in approximating mixture of normals m_c is to apply Bayes theorem to the density given in equation (A.18) and draw from the resulting posterior density directly. But sampling θ_ℓ from its posterior distribution in this model can be eased by introducing the configuration vector $S = \{S_1, \dots, S_L\}$. That is, $S_\ell = j$ if and only if $\theta_\ell = \xi_j$. With this setting sampling θ_ℓ is equivalent to sampling S and ξ_j , $j = 1, 2, \dots, m_c$. West, Müller and Escobar (1994) propose an efficient algorithm as the following:

1. Sampling S_ℓ , $\ell = 1, 2, \dots, L$ from the conditional distribution

$$q(S_\ell = j|\xi^{(\ell)}, P_0) \propto \begin{cases} \tau q_t(e_{new, \ell}|\mu_0, (1 + \tau_0)R_0/n_0, n_0) & \text{if } j = 0 \\ n_j^{(\ell)} \phi(e_{new, \ell}|\xi_j^{(\ell)}) & \text{if } j > 0 \end{cases} \quad (\text{A.21})$$

for $j = 1, 2, \dots, m_c$, where $q_t(e_{new,\ell}|\mu_0, (1 + \tau_0)R_0/n_0, n_0)$ is the density of the t -distribution with mean μ_0 , scale factor $(1 + \tau_0)R_0/n_0$ and n_0 degrees of freedom and $\phi(e_{new}|\xi_j^{(\ell)})$ denotes the normal probability density characterized by $\xi_j^{(\ell)}$. If $S_\ell = 0$, then draw a new θ_ℓ from $q(\theta_\ell|e_{new,\ell}) \propto q(e_{new,\ell}|\theta_\ell)dP_0$.

2. Sampling ξ_j , $j = 1, 2, \dots, m_c$ from the conditional distribution

$$\begin{aligned} q(\xi_j|\xi^{(\ell)}, S, P_0) &\propto \prod_{\{e_{new,\ell}: S_\ell=j\}} \phi(e_{new,\ell}|\xi_j)dP_0 \\ &\propto Normal(\mu_1, \tau_1\sigma_j^2)IG\left(\frac{N_3}{2}, \frac{R_3}{2}\right) \end{aligned} \quad (\text{A.22})$$

where

$$\mu_1 = \frac{\tau_0 \sum_{\{e_{new,\ell}: S_\ell=j\}} e_{new,\ell} + \mu_0}{\tau_0 n_j + 1}, \tau_1 = \frac{\tau_0}{\tau_0 n_j + 1}, N_3 = N_0 + n_j, \quad (\text{A.23})$$

$$R_3 = R_0 + \frac{n_j \left(\frac{1}{n_j} \sum_{\{e_{new,\ell}: S_\ell=j\}} e_{new,\ell} - \mu_0 \right)^2}{\tau_0 n_j + 1} + \sum_{\{e_{new,\ell}: S_\ell=j\}} \left(e_{new,\ell} - \frac{1}{n_j} \sum_{\{e_{new,\ell}: S_\ell=j\}} e_{new,\ell} \right)^2 \quad (\text{A.24})$$

and n_j is the number of observations such that $S_\ell = j$.

Finally, Escobar and West (1995) show that with a beta distributed variable $\eta \sim Beta(\tau + 1, L)$, the full conditional distribution of τ is given by

$$\tau \sim z \times \text{gamma}(d_1 + m_c, d_2 - \log \eta) + (1 - z) \times \text{gamma}(d_1 + m_c - 1, d_2 - \log \eta) \quad (\text{A.25})$$

where $\frac{z}{1-z} = \frac{d_1 + m_c - 1}{L(d_2 - \log \eta)}$.

Thus, with a set of $e_{new,\ell}$, we updated θ_ℓ for each ℓ , the number of components in the countable infinite mixture of normals m_c and τ , the precision parameter in the base measure.

B. Details of the MCMC Algorithm.

1. **a** Draw $(\mu_{1,\ell}, q_\ell^*)$. Since the posterior for $(\mu_{1,\ell}, q_\ell^*)$ involves the computation of the inverse bidding function for each bid, using any tailored proposal density would be too time-consuming. Also, the q_ℓ^* can only take values between $(0, 1)$. With these features, we

propose a simple random walk proposal density as the following:

$$q(\mu_{1,\ell}^{new}, q_\ell^{*new} | \mu_{1,\ell}^{old}, q_\ell^{*old}) = f_t(\mu_{1,\ell}^{new} | \mu_{1,\ell}^{old}, h_0, v_0)$$

where $f_t(\cdot | \mu_{1,\ell}^{old}, h_0, v_0)$ is a student t distribution with mean $\mu_{1,\ell}^{old}$, variance h_0 and degree of freedom v_0 . h_0 and v_0 are tuning parameters to obtain reasonable acceptance rates. In practice, to draw from this proposal density, we draw q_ℓ^{*new} from the uniform distribution and $\mu_{1,\ell}^{new}$ from $f_t(\cdot | \mu_{1,\ell}^{old}, h_0, v_0)$ independently. Also, this proposal density is symmetric in $(\mu_{1,\ell}^{old}, q_\ell^{*old})$ and $(\mu_{1,\ell}^{new}, q_\ell^{*new})$. When a set of new values $(\mu_{1,\ell}^{new}, q_\ell^{*new})$ is drawn, the chain moves to the proposal value with probability

$$\begin{aligned} & \Pr \left[\left(\mu_{1,\ell}^{old}, q_\ell^{*old} \right), \left(\mu_{1,\ell}^{new}, q_\ell^{*new} \right) \right] \\ &= \min \left\{ \frac{\pi(\mu_{1,\ell}^{new}, q_\ell^{*new} | b, n, \boldsymbol{\beta}, \Theta, d_\ell, \sigma_\ell^2)}{\pi(\mu_{1,\ell}^{old}, q_\ell^{*old} | b, n, \boldsymbol{\beta}, \Theta, d_\ell, \sigma_\ell^2)}, 1 \right\}. \end{aligned}$$

If the candidate is not accepted then the chain does not change its value.

- b** Draw $\Theta = (\boldsymbol{\delta}^*, \gamma^*)$. Denote Θ_{old} the current state of the Markov chain, $\widehat{\Theta}$ the mode of the full conditional density and Θ_{new} the candidate for the new value of the chain. Following Chib et al. (1998) the proposal density $q(\Theta_{old}, \Theta_{new})$ is selected to be multivariate t -distribution with v_1 degrees of freedom $f_T(\Theta_{new} | \widehat{\Theta} - (\Theta_{old} - \widehat{\Theta}), \tau_2 V)$ centered at $\widehat{\Theta} - (\Theta_{old} - \widehat{\Theta})$. This proposal density is symmetric in Θ_{old} and Θ_{new} . The covariance matrix of the multivariate t -density is set to be $V = -H_{\widehat{\Theta}}^{-1}$, negative inverse of the Hessian of $\log \pi[\Theta | \mu_{1,\ell}, q_\ell^*]$ evaluated at the modal value $\widehat{\Theta}$ and τ_2 is an adjustable constant. The candidate Θ_{new} , drawn from the proposal distribution, is accepted with probability

$$\Pr[\Theta_{old}, \Theta_{new}] = \min \left(\frac{\pi[\Theta_{new} | \mu_{1,\ell}, q_\ell^*]}{\pi[\Theta_{old} | \mu_{1,\ell}, q_\ell^*]}, 1 \right).$$

If the candidate is not accepted then the chain does not change its value. The modal value $\widehat{\Theta}$ is found using the Newton-Raphson method. The gradient is

$$g_\Theta = -(\Theta - \theta_0)' D_0 + \sum_{\ell=1}^L \mathbf{x}_\ell^* - \exp(-\mu_{1,\ell} + \mathbf{x}_\ell^* \Theta) k_\ell \mathbf{x}_\ell^*$$

and the Hessian is

$$H_\Theta = -D_0 - \exp(-\mu_{1,\ell} + \mathbf{x}_\ell^* \Theta) k_\ell \mathbf{x}_\ell^{*'} \mathbf{x}_\ell^*.$$

REFERENCES

- Athey, S. and P. Haile (2002): “Identification of Standard Auction Models,” *Econometrica*, 70, 6, 2107-2140.
- (2004): “Nonparametric Approaches to Auctions,” in *Handbook of Econometrics*, Volume 6, forthcoming.
- Athey, S., J. Levin and E. Seira (2004): “Comparing Open and Sealed Bid Auctions: Theory and Evidence from Timber Auction,” Working paper, Stanford University.
- Bajari, P. (1997): “The First Price Auction with Asymmetric Bidders: Theory and Applications,” Ph.D. Dissertation, University of Minnesota.
- Bajari, P. and A. Hortacısu (2003): “The Winner’s Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions,” *RAND Journal of Economics*, 34, 2, 329-355.
- Bajari, P. and L. Ye (2002): “Competition versus Collusion in Procurement Auctions: Identification and Testing,” Working paper, Stanford University.
- Berry, S. (1992): “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60, 889-917.
- Bresnahan, T. and P. Reiss (1990): “Entry in Monopoly Markets,” *Review of Economic Studies*, 57, 531-553.
- Bulow, J. and P. Klemperer (1996): “Auctions Versus Negotiations,” *American Economic Review*, 86, 180-194.
- Bulow, J. and P. Klemperer (2002): “Prices and the Winner’s Curse,” *RAND Journal of Economics*, 33, 1-21.
- Chernozhukov, V. and H. Hong (2004): “Likelihood Estimation and Inference in a Class of Non-regular Econometric Models,” *Econometrica*, 72, 1445-1480.
- Chib, S. and B. H. Hamilton (2002): “Semiparametric Bayes Analysis of Longitudinal Data Treatment Models,” *Journal of Econometrics*, 110, 67-89.

- Chib, S., E. Greenberg and R. Winkelmann (1998): "Posterior Simulation and Bayes Factor in Panel Count Data Models," *Journal of Econometrics*, 86, 33-54.
- Donald, S. and H. Paarsch (1993): "Piecewise Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, 34, 121-148.
- (1996): "Identification, Estimation, and Testing in Parametric Empirical Models of Auctions within the Independent Private Value Paradigm," *Econometric Theory*, 12, 512-567.
- Escobar, M.D. (1994): "Estimating Normal Means with a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268-277.
- Escobar, M.D., and M. West (1995): "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577-588.
- Ferguson, T. S. (1973): "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209-230.
- (1974): "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615-629.
- (1983): "Bayesian Density Estimation By Mixtures of Normal Distributions," in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*, ed. by H. Rizvi and J. Rustagi. New York:: Academic Press, pp. 287-302.
- Geweke, J. and M. Keane (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics*, Volume 5, ed. By J. Heckman and E. Leamer. Amsterdam: North-Holland, 3462-3571.
- Ghosal, S., J. K. Ghosh and R. V. Ramamoorthi (1999): "Posterior Consistency of Dirichlet Mixtures in Density Estimation," *Annals of Statistics*, 27, 143-158.
- Griffin, J.E. and M.F.J. Steel (2004): "Semiparametric Bayesian Inference for Stochastic Frontier Models," *Journal of Econometrics*, forthcoming.
- Guerre, E., I. Perrigne and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525-574.

- Haile, P. (2001): "Auctions with Resale Markets: An Application to U.S. Forest Service Timber Sales," *American Economic Review*, 92, 3, 399-427.
- Haile, P., H. Hong and M. Shum (2002): "Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions," Working paper, Princeton University.
- Harstad, R. M. (1990): "Alternative Common-Value Auction Procedures: Revenue Comparisons with Free Entry," *Journal of Political Economy*, 98, 421-429.
- Harstad, R. M., D. Levin and J. Kagel (1990): "Equilibrium Bid Function for Auctions with Uncertain Number of Bidders," *Economics Letters*, 33, 35-40.
- Hasegawa, H. and H. Kozumi (2003): "Estimation of Lorenz Curves: A Bayesian Nonparametric Approach," *Journal of Econometrics*, 115, 277-291.
- Hastings, W.K. (1970): "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97-109.
- Heckman, J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109, 4, 673-748.
- Hendricks, K., J. Pinkse, and R. Porter (2003): "Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions," *Review of Economic Studies*, 70, 115-145.
- Hirano, K. (2002): "Semiparametric Bayesian Inference in Autoregressive Panel Data Models," *Econometrica*, 70, 2, 781-799.
- Hirano, K. and J. R. Porter (2003): "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," *Econometrica*, 71, 1307-1338.
- Hong, H. and M. Shum (2002): "Increasing Competition and the Winner's Curse: Evidence from Procurement," *Review of Economic Studies*, 69, 871-898.
- (2003): "Econometric Models of Asymmetric Ascending Auctions," *Journal of Econometrics*, 112, 327-358.
- Jofre-Bonet, M. and M. Pesendorfer (2003), "Estimation of a Dynamic Auction Game," *Econometrica*, 71, 5, 1443-1489.

- Kjerstad, E. and S. Vagstad (2000): “Procurement Auctions with Entry of Bidders,” *International Journal of Industrial Organization*, 18, 1243-1257.
- Krasnokutskaya, E. (2002): “Identification and Estimation in Highway Procurement Auctions under Unobserved Auction Heterogeneity,” Working paper, University of Pennsylvania.
- Krasnokutskaya, E. and K. Seim (2006): “Bid Preference Programs and Participation Highway Procurement Auctions,” Working Paper, University of Pennsylvania.
- Krishna, V. (2002): *Auction Theory*, Academic Press, San Diego.
- Laffont, J-J., H. Ossard and Q. Vuong (1995): “Econometrics of First-Price Auctions,” *Econometrica*, 63, 953-980.
- Levin, D. and E. Ozdenoren (2004): “Auctions with Uncertain Number of Bidders,” *Journal of Economic Theory*, 118, 229-251.
- Levin, D. and J. L. Smith (1994): “Equilibrium in Auctions with Entry,” *American Economic Review*, 84, 3, 585-599.
- Li, T. (2005): “Econometrics of First-Price Auctions with Entry and Binding Reservation Prices,” *Journal of Econometrics*, 126, 173-200.
- 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 Value Model,” *RAND Journal of Economics*, 33(2), 171-193.
- Lo, A. Y. (1984): “On a Class of Bayesian Nonparametric Estimates: I. Density Estimates,” *The Annals of Statistics*, 12, 351-357.
- Matthew, S. (1987): “Comparing Auctions for Risk Averse Buyers: A Buyer’s Point of View,” *Econometrica*, 55, 633-646.
- McAfee, R. P. and J. McMillan (1987a): “Auctions with Entry,” *Economics Letters*, 23, 343-347.

- McAfee, R. P. and J. McMillan (1987b): “Auctions with Stochastic Number of Bidders,” *Journal of Economic Theory*, 43, 1-19.
- McAfee, R. P. , D. Quan and D. Vincent (2002): “How to Set Minimum Acceptable Bids, with an Application to Real Estate Auctions,” *Journal of Industrial Economics*, 50(4), 391-416.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller (1953): “Equations of State Calculations by Fast Computing Machines,” *Journal of Chemical Physics*, 21, 1087-1092.
- Paarsch, H. (1992): “Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions,” *Journal of Econometrics*, 51, 191-215.
- (1997): “Deriving an Estimate of the Optimal Reserve Price: An Application to British Columbian Timber Sales,” *Journal of Econometrics*, 78, 333-357.
- Pinkse, J. and G. Tan (2004): “The Affiliation Effect in First-Price Auctions,” *Econometrica*, forthcoming.
- Porter, R. and D. Zona (1993): “Detection of Bid Rigging in Procurement Auctions,” *Journal of Political Economy*, 101, 518-538.
- Samuelson, W. F. (1985): “Competitive Bidding with Entry Costs,” *Economics Letters*, 17(1-2), 53-57.
- Sareen, S. (1999): “Posterior Odds Comparison of a Symmetric Low-Price, Sealed-Bid Auction within the Common Value and the Independent Private Value Paradigms,” *Journal of Applied Econometrics*, 14, 651-676.
- Sethuraman, J. (1994): “A Constructive Definition of Dirichlet Priors,” *Statistica Sinica*, 4, 639-650.
- Tanner, M.A. and W.H. Wong (1987): “The Calculation of Posterior Distribution by Data Augmentation,” *Journal of American Statistical Association*, 82, 528-540.

West, M., P. Müller, and M. D. Escobar (1994): “Hierarchical Priors and Mixture Models, with Application in Regression and Density Estimation,” in *Aspects of Uncertainty. A Tribute to D.V. Lindley*, ed. by A.F.M. Smith and P.R. Freeman. London: Wiley, pp. 363-386.

Table 1 Summary Statistics²²

Variable	Explanation	Obs	Mean	Standard Deviation
Estimate	Engineer's estimate	553	165348.90	130782.30
Bids	Bids	1606	165382.20	133721.70
Day	Number of working days	553	122.72	120.07
Full	Acreage of full width mowing	553	7302.94	5219.91
Other	Acreage of other mowing	553	1987.64	2803.50
Items	Number of items	553	2.01	0.82
State	1 if it is a state job	553	0.11	0.31
Interstate	1 if it is on an interstate highway	553	0.23	0.42
Potential	Number of potential bidders	553	11.08	3.47
Actual	Number of actual bidders	553	2.90	1.17
Entry	Actual/Potential	553	0.28	0.12
Highbids	1 if the bid is higher than Estimate	1606	0.44	0.50
Highwinner	1 if the winning bid is higher than Estimate	553	0.22	0.41

Table 2: Quasi-MLE Poisson Regression Estimates of Number of Actual Bidders on Explanatory

Variables ²³			
Variable	Estimate	Sd. Error	<i>t</i> -statistic
log(Estimate)	-0.2004	0.0472	-0.42
log(Day)	0.0543	0.0386	1.41
log(Full)	0.0127	0.0078	1.64
log(Other)	0.0230*	0.0063	3.63
Items	-0.1847*	0.0324	-5.69
State	-0.0581	0.0573	-1.01
Interstate	-0.0305	0.0388	-0.79
log(Potential)	0.4011*	0.0454	8.84
Constant	0.2698	0.4296	0.63
Log likelihood	-899.28		

²²The sample consists of 553 auctions with a total of 1606 bids.²³*: significant at 5%.

Table 3: OLS Estimates of Log(bids) on Explanatory Variables²⁴

Variable	Estimate	Sd. Error	<i>t</i> -statistic
log(Estimate)	1.0354*	0.0124	83.51
log(Day)	-0.0301*	0.0087	-3.46
log(Full)	-0.0021	0.0031	-0.67
log(Other)	-0.0079*	0.0017	-4.77
Items	0.0263*	0.0094	2.80
State	-0.0108	0.0182	-0.59
Interstate	0.0199	0.0111	1.80
Potential	-0.0025	0.0014	-1.80
Constant	-0.2514*	0.1166	-2.16
Adjusted R^2	0.9324		

Table 4: Random Effects Panel Data Estimates of Log(bids) on Explanatory Variables²⁵

Variable	Estimate	Sd. Error	<i>t</i> -statistic
log(Estimate)	1.0265*	0.0168	61.20
log(Day)	-0.0269*	0.0121	-2.23
log(Full)	-0.0014	0.0040	-0.35
log(Other)	-0.0074*	0.0022	-3.38
Items	0.0279*	0.0119	2.34
State	0.0003	0.0240	0.01
Interstate	0.0201	0.0148	1.36
Potential	-0.0021	0.0018	-1.14
Constant	-0.1747	0.1568	-1.11
σ_u^2	0.1085	σ_e^2	0.1444
ρ (fraction of variance due to u_i)	0.3608	Adjusted R^2	0.9327

²⁴*: significant at 5%.²⁵Model: $\log(\text{bids}_{it}) = \mathbf{x}_{it}\boldsymbol{\lambda} + u_i + \varepsilon_{it}$ *: significant at 5%.

Table 5: OLS Estimates of Log(winning bid) on Explanatory Variables²⁶

Variable	Estimate	Sd. Error	<i>t</i> -statistic
log(Estimate)	1.0408*	0.0183	56.86
log(Day)	-0.0303*	0.0130	-2.33
log(Full)	0.0022	0.0053	0.41
log(Other)	-0.0108*	0.0030	-3.66
Items	0.0213	0.0121	1.76
State	0.0058	0.0253	0.23
Interstate	0.0235	0.0156	1.51
Potential	-0.0035	0.0020	-1.75
Constant	-0.4107*	0.1655	-2.48
Adjusted R^2	0.9544		

Table 6: Estimation Results From Semiparametric Bayesian Algorithm²⁷

Variable	Mean	Standard Deviation	Mean	Standard Deviation
	Private Cost Distribution		Entry Cost Distribution	
log(Estimate)	0.9425	0.0364	0.8330	0.0514
log(Day)	0.0004	0.0523	0.0336	0.0771
log(Full)	-0.0084	0.0218	-0.0243	0.0306
log(Other)	-0.0103	0.0121	-0.0286	0.0177
Items	0.0583	0.0526	0.1817	0.0747
State	0.0778	0.0999	0.2113	0.1385
Interstate	0.0409	0.0672	0.0716	0.0946
Difference in Constants	0.1833 (Mean)		0.2357 (Stan. Dev.)	

²⁶ *: significant at 5%.²⁷ Results are based on 15,000 draws following a 5,000 initial burn-in period.

Figure 1: Example of Non-monotone Relationship Between b and N

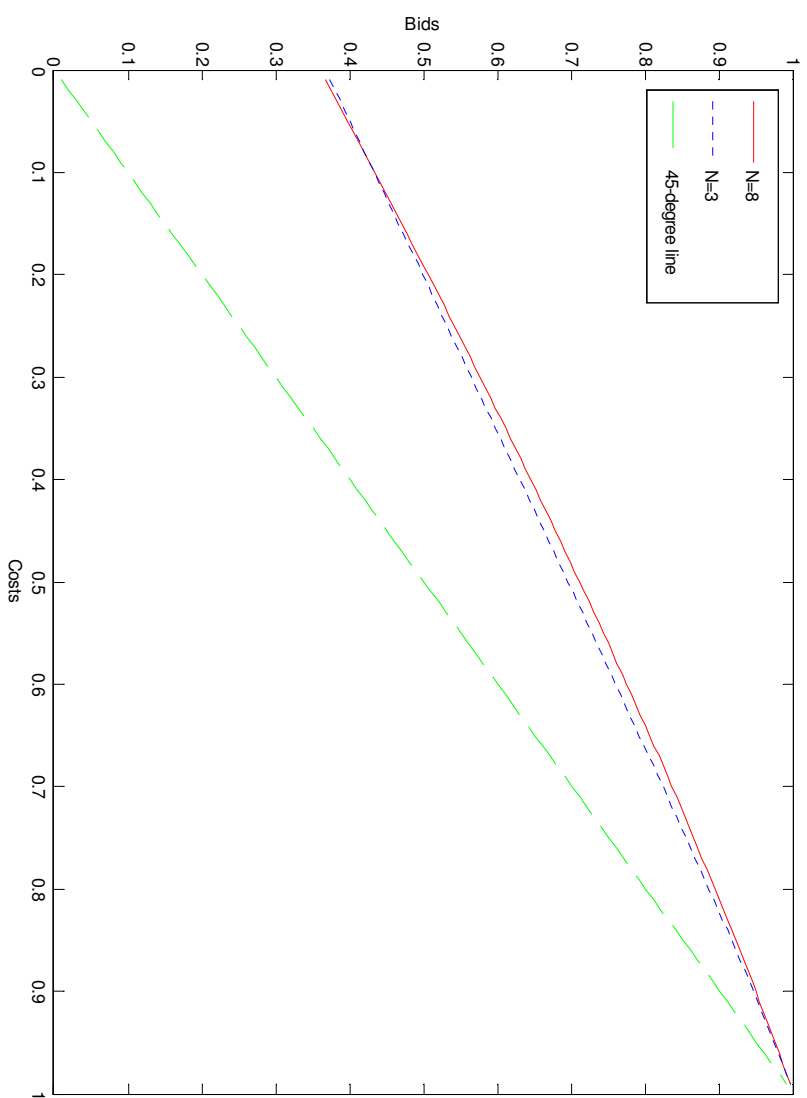


Figure 2: Time Series, Histograms and Autocorrelations of the MCMC Draws for the Bidding Equation

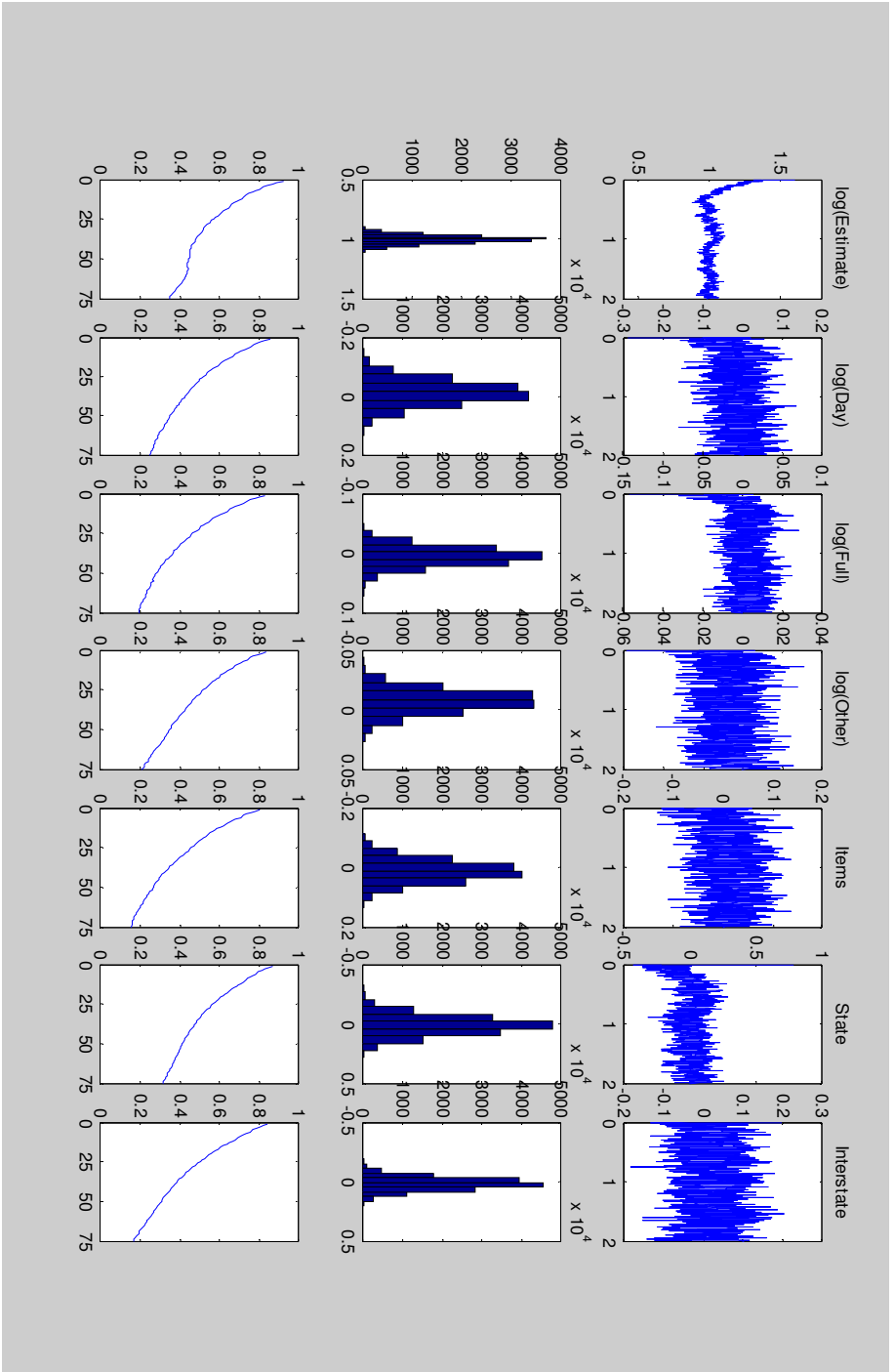


Figure 3: Time Series, Histograms and Autocorrelograms of the MCMC Draws for the Entry Equation

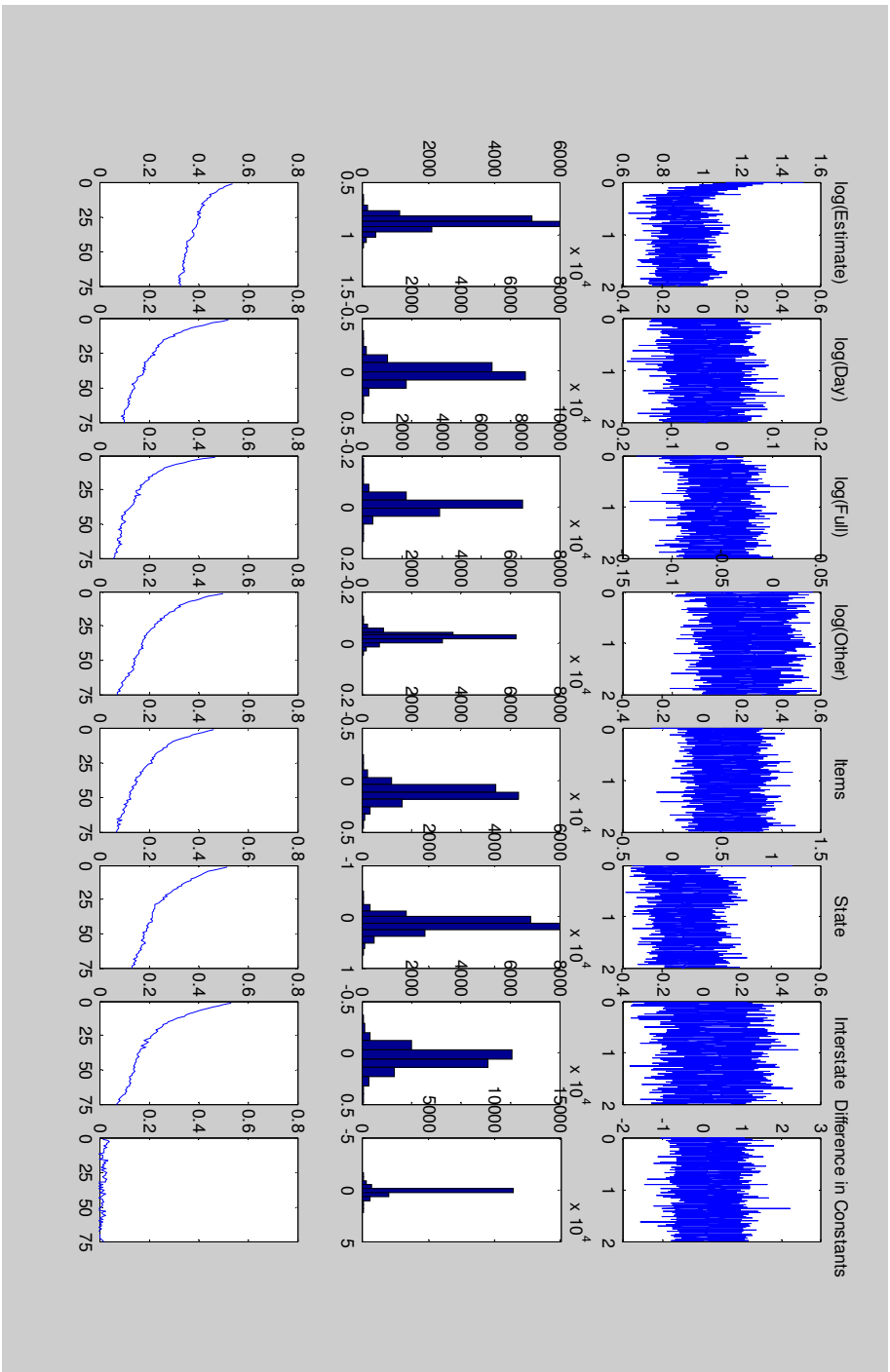


Figure 4: Estimated Predictive Density for ϵ_{new}

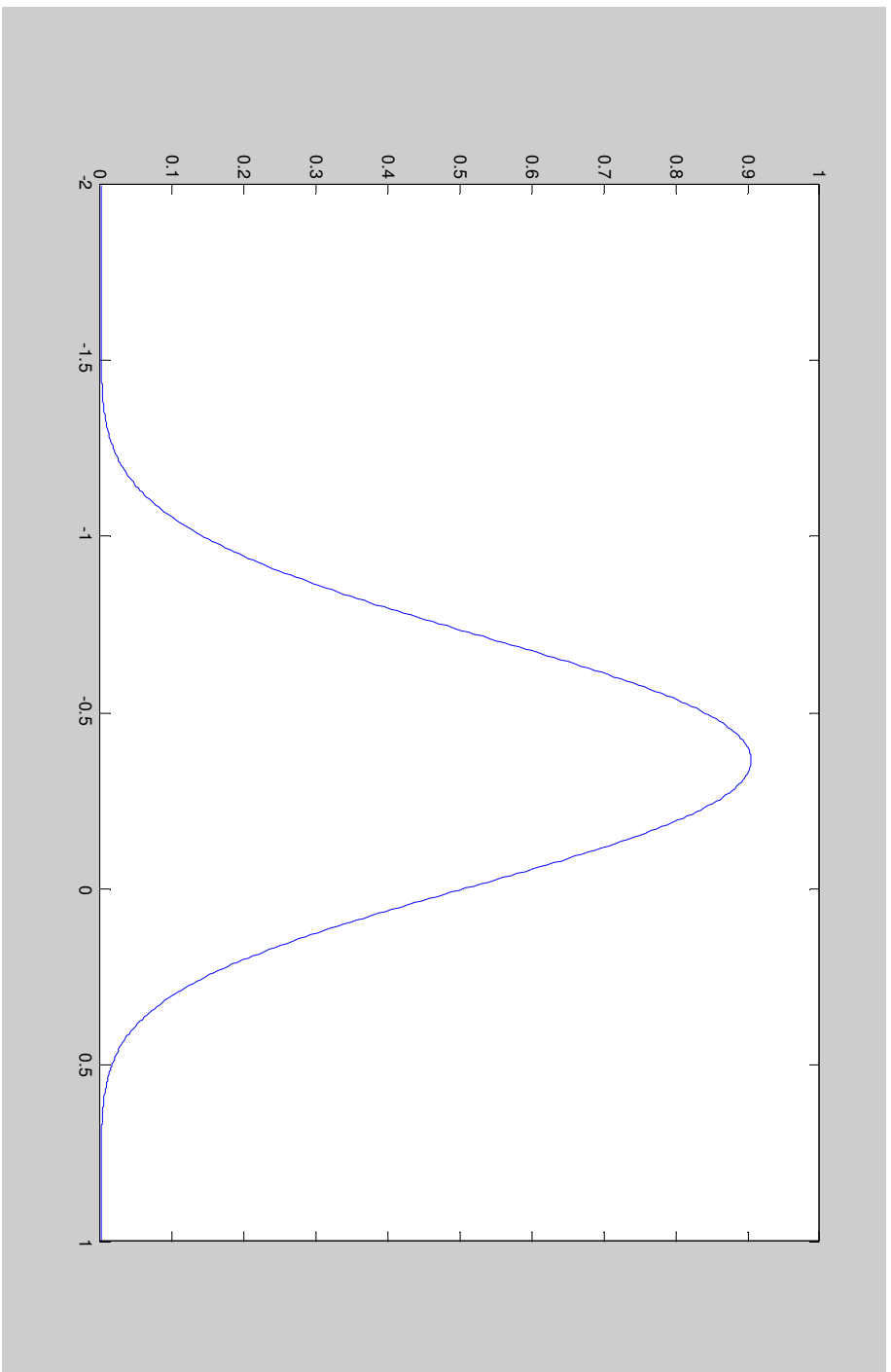


Figure 5: Simulated Versus Actual Procurement Costs

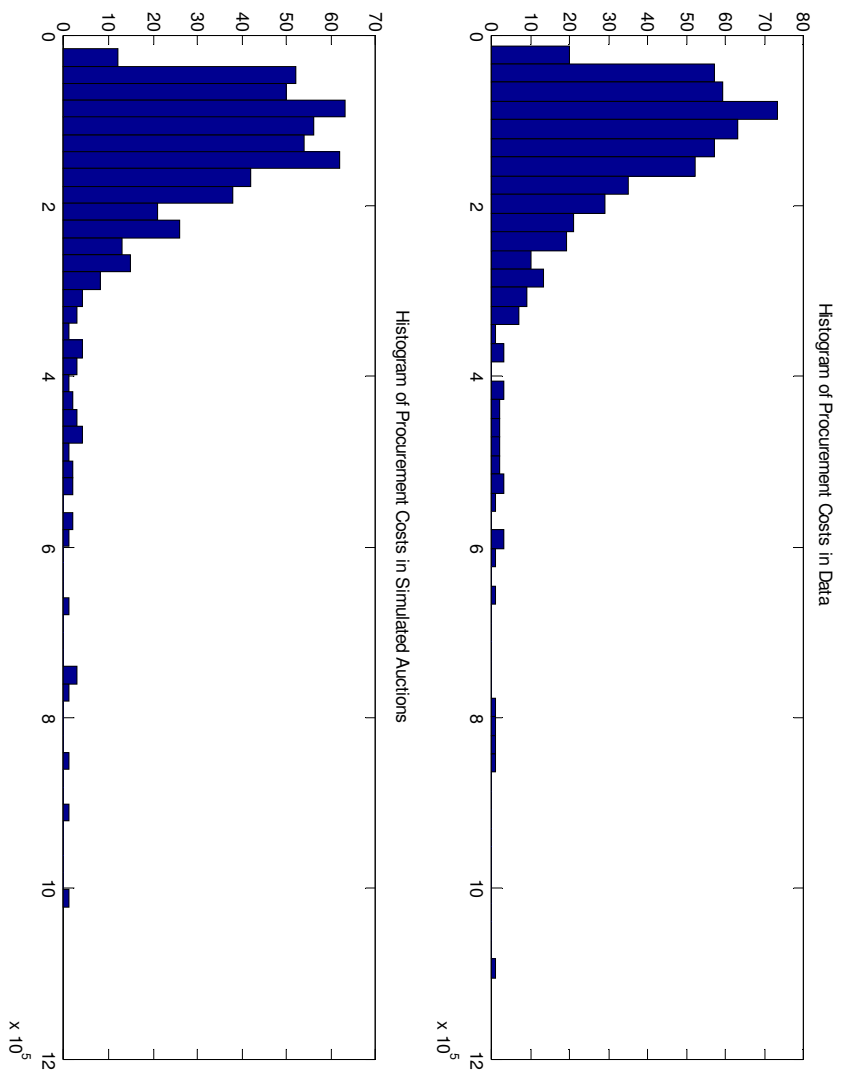


Figure 6: The “Competition Effect”, “Entry Effect” and “Total Effect” from N on b

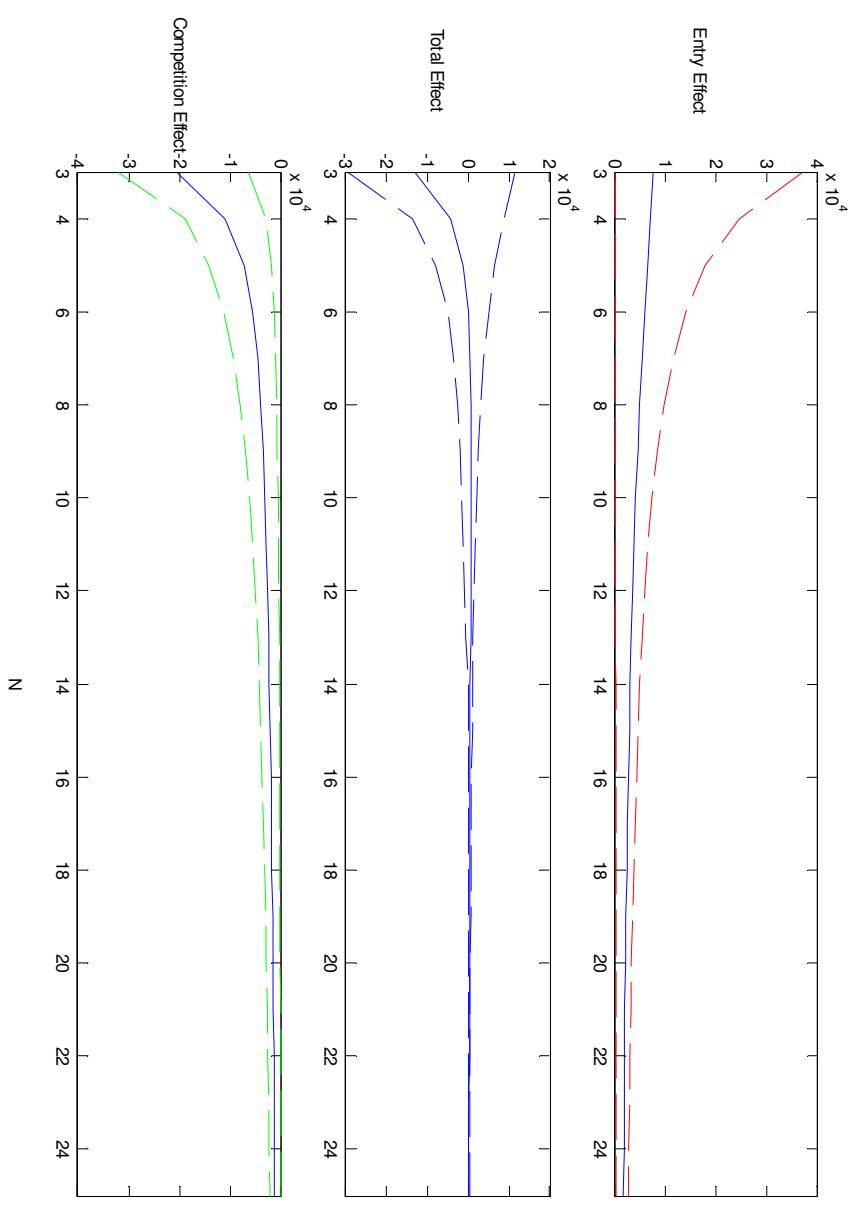


Figure 7: The “Competition Effect,” “Entry Effect” and “Total Effect” from N on Procurement Cost

